Miriam Khider

Cloning and Characterization of Three Transketolase Genes (*tktA1*, *tktA2* and *tktA3*) from *Azotobacter vinelandii*

Master Thesis

Trondheim, November 2010

The Norwegian University of Science and Technology
The Faculty of Natural Sciences and Technology
Department of Biotechnology

Supervisors: Professor Svein Valla and Dr. Scient Helga Ertesvåg

NTNU
Norwegian University of
Science and Technology

Acknowledgements

Abstract

Alginate is synthesized by the gram-negative bacteria *Azotobacter vinelandii*. Alginate appears as an important component in both active vegetative cells and in cysts, which is one of the bacteriums life cycle forms.
Transketolase is an essential enzyme in the non-oxidative branch of the pentose phosphate pathway. Transketolase is responsible for the conversion of D-erythrose 4-phosphate to D-fructose 6-phosphate which is the starting precursor in alginate biosynthesis.
Mutant and wild-type transketolase have been widely studied in *Escherichia coli*, this work focused on studying three transketolase genes from *A. vinelandii.*

In this study, three transketolase genes (*tktA1, tktA2 and tktA3)* from *A. vinelandii* were mutated by in frame deletions of about 1200 bp and each were cloned into a suicide vector containing the spectinomycin resistance gene and the *sacB* gene of *Bacillus subtilis* as selective markers.
The *tktA1* and *tktA2* genes were also individually cloned into a transposon vector under control of the *Pm*-promoter. A previous master student had constructed a transposon containing *tktA3*.

The three chromosomal copies of genes were to be inactivated by homologous gene replacement in *A. vinelandii* ATCC12518. These deletion mutants should then be complemented by the transposon encoded genes. However, even after several experiments no deletion mutants were obtained.

The second part of this work focused on studying the three wild-type transketolase genes (*tktA1, tktA2 and tktA3)* of *A. vinelandii* by cloning, identification, purification and activity analyses in *E. coli.*
The open reading frames of the three transketolase genes *tktA1, tktA2 and tktA3* were cloned and overexpressed from the *Pm*-promoter in an RK2 based expression vector *E. coli S17.1 λ pir* were used as host since transformants of *RV308* containing these plasmids could not be obtained.

Thus, the three proteins were produced in *E. coli S17.1 λ pir*. No protein activity was found from any of three expression strains. The SDS-PAGE of sonicated cells showed bands of 74.6 kDa and 73.8 kDa corresponding *tktA2* and *tktA3*. The three transketolase proteins were purified from sonicated cells using FPLC and the obtained fraction were analysed by SDS-PAGE. Protein bands with the expected molecular weight for Tkt2 and Tkt3 were obtained. When these extracts were purified by FPLC, this protein band was only found in the void fraction. The TktA1 protein band was not observed neither using the sonicated cells nor in the fractions obtained from the FPLC.

Table of Contents

1 Introduction

1.1 *Azotobacter vinelandii*

1.1.1 Characteristics, Properties and Genome

Azotobacter vinelandii belongs to *Pseudomonadaceae* family (Setubal *et al.*, 2009) and is 2 µm in diameter and 4 µm in length (Kennedy et al., 2005) . The bacteria are large; motile with flagella, gram- negative, obligate aerobe with high metabolic activity. They are also free-living bacteria able to grow in nitrogen free medium (e.g. Burk's) with generation time of 2.5-3h at 30°C (Sadoff, 1975). The cells are usually single but can occur in pairs. *A. vinelandii* has a high DNA content; the chromosome can reach a high copy number in the late exponential phase (Kennedy et al., 2005).

A. vinelandii is able to produce alginate, and two other industrially important secondary metabolites, poly- β-hydroxybutyric acid (PHB) and 5-alkylresorcios (phenolic lipid) (Sabra *et al.*, 2001). The alginate protects the organism from excess exogenous oxygen. The bacteria encode several proteins that alter the composition of alginate in response to oxygen availability. Some alginate lyase (at least six are encoded by *A. vinelandii*) may be responsible for altering alginate composition in response to oxygen (Setubal et al., 2009). At high oxygen the L-guluronic acid content of alginate increases. This process requires the secreted mannuronan C-5 epimerases (AlgE1-7) (Sabra *et al.*, 2000).
Nitrogen fixation is an oxygen-sensitive process where the nitrogenase can be irreversibly damaged by exposure to air (Robson, 1985).
A. vinelandii and several other members of the *Azotobacter* genus are commonly found in soil and are capable of fixing nitrogen from nitrogen free medium in air and in the presence of carbon source (Robson, 1985, Clementi, 1997). Two aspects make *A. vinelandii* interesting for investigation and study of its nitrogen fixation process: the first is that this bacterium encodes two additional oxygen-sensitive nitrogenase (iron-only nitrogenase encoded by *anf* and vanadium nitrogenase encoded by *vnf*), besides the molybdenum-containing nitrogenase, (encoded by *nif*) and the second is the ability to fix nitrogen aerobically, which is rare among bacteria (Setubal et al., 2009, Sabra et al., 2001). As stated previously that *A. vinelandii* is an obligate aerobe and thus requires oxygen to carry out its biological processes. Eventhough the bacteria are able to overcome this problem and to fix nitrogen even in the absence of oxygen. This is firstly because *A. vinelandii* has one of the highest respiratory rates among living organisms which enables it to produce high energy content supplying nitrogenase and secondly its high oxygen consumption (Steigedal, 2006, Robson, 1985).
The nitrogenase in *A. vinelandii* is protected by regulating low cytoplasmatic oxygen through a process known as respiratory-protection where high respiratory activity that removes oxygen at the cell surface and preventing its diffusion into the cell. The respiratory-protection hypothesis was proposed by Dalton and Postgate in 1969.
Later in 2000, experimental results showed that under nitrogen fixation conditions, the *A. vinelandii* forms alginate capsula under high pO_2 (dissolved oxygen concentration). Thus this may serve as a protection for nitrogenase from oxygen damage where alginate quality is essential for this protection (Sabra et al., 2001).

A. vinelandii has a single circular genome of 5.365.318 bp with a GC content of 65.5% (Setubal et al., 2009). The complete genome sequenced of *A. vinelandii* is available on Azotobacter.org site. The availability of *A. vinelandii* DJ complete genome makes it one of

the model organisms for studying alginate and it is proposed as a host for the production and characterization of oxygen sensitive proteins (Setubal et al., 2009).

Several physical, biochemical and spectrophotometrical studies suggested that *A. vinelandii* containing a multiple copy is (40-80) of its chromosome per cell. The new *in vivo* chromosome counting techniques published by Lakshmi Pulakat in 2002 concluded that *A. vinelandii* genome may vary in copy number depending on the growth medium (Pulakat et al., 2002).

1.1.2 Metabolism

1.1.2.1 Pentose Phosphate Pathway

The pentose phosphate pathway (PPP) also known as the phosphogluconate pathway (Zhao & Zhong, 2009) or hexose monophosphate shunt (Wamelink et al., 2008), plays a key role in metabolism through its intermediates; *ribose 5 phosphate (R5P)*, *erythrose 4 phosphates (E4P) and sedoheptulose 7-phosphate (S7P)*. Where R5P is the starting precursors for the biosynthesis of purine and pyrimidine nucleotides and histidine, E4P is the precursor for aromatic amino acids, aromatic vitamins and pyridoxin biosynthesis and S7P is the precursor for the biosynthesis of the gram negative bacteria cell wall components (Sprenger *et al.*, 1995, Kamada *et al.*, 2001, Iida *et al.*, 1993).

The PPP is composed of non-oxidative and oxidative branches (Figure 1.1 A and B).The oxidative non reversible branch is the major source of the reductive power, NADPH which is used in many biosynthetic reactions. This branch also provides intermediate for biosynthetic processes supplying the cell with R5P. The oxidative branch of the PPP has three irreversible steps where NADP reduced to NADPH at the same time glucose-6-phosphate (G6P) is converted to pentose phosphate and CO_2 (Wamelink et al., 2008) (Figure 1.1 A).

The non-oxidative reversible branch connects the pentose phosphate to glycolytic intermediates (Wamelink et al., 2008), through the reconversion of the pentose phosphate to G6P and the two other intermediates fructose-6-phosphate and Glyceraldehyde 3 phosphate Figure 1.1 A). *A. vinelandii* lacks the *phosphoglucose isomerase* and thus can not metabolise glucose through the glycolytic pathway. The bacteria use an Entner-Doudoroff pathway to obtain its precursors.

In the non-oxidative branch the riboluse 5P (formed in the oxidative branch) is isomerised to R5P by ribulose-5-phosphate isomerase (RPI) or epimerized to xylulose 5 phospahte by ribulose 5 phosphate epimerase (RPE) (Figure 1.1 B). Transketolase converts R5P and X5P into GA3P and S7P, where both are metabolized by tranaldolase (Wamelink et al., 2008).
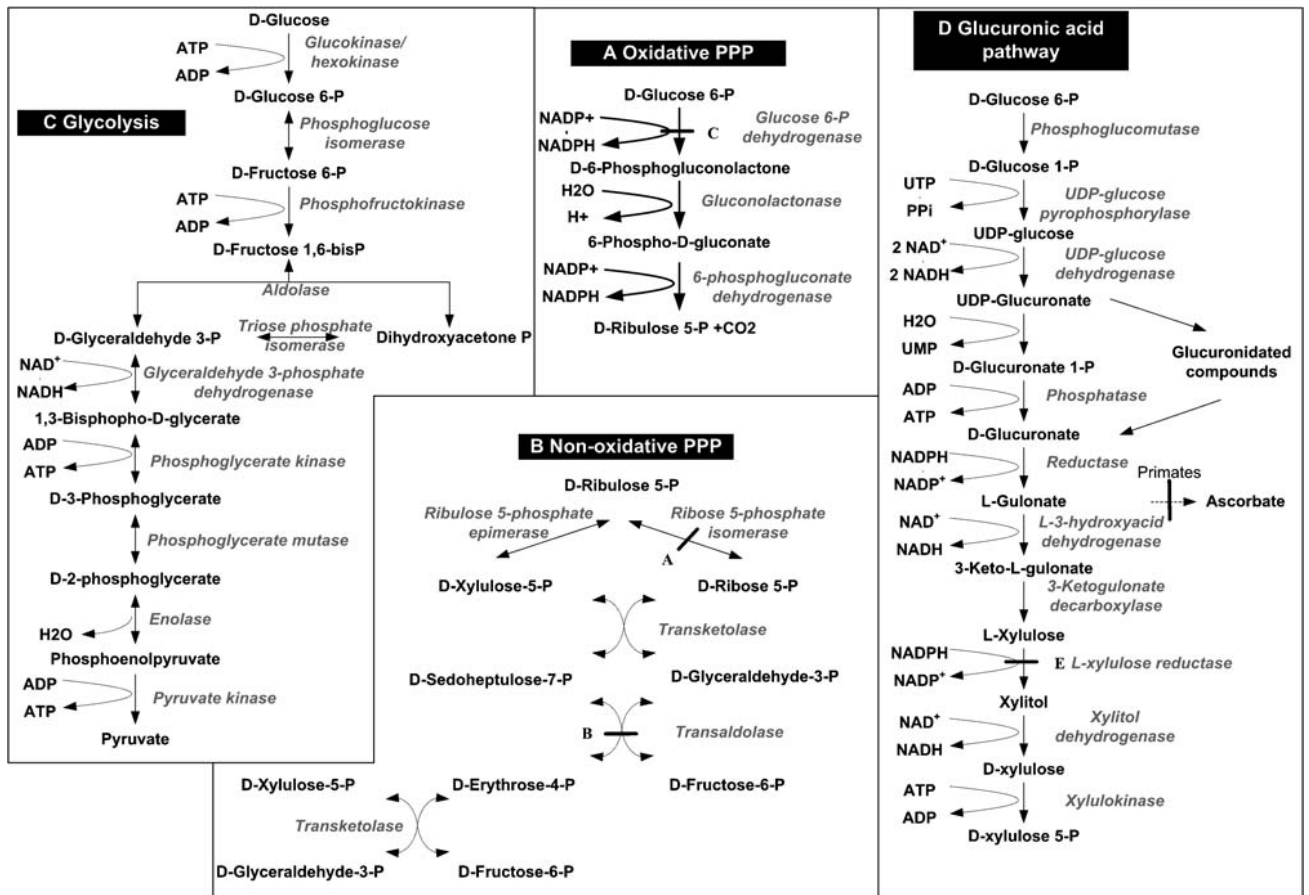
Figure 1.1 Scheme of pentose phosphate pathway and connected pathways (Wamelink et al., 2008). ***A-****The oxidative PPP,* ***B-****Non-oxidative PPP,* ***C-****Glycolysis,* ***D-****Glucuronic acid pathway.*

1.1.2.2 Transketolase

The name transketolase was suggested by E. Racker (Racker *et al.*, 1953) and it is one of the thiamine diphosphate dependent (ThDP) enzymes, which catalyzes reactions in both the reductive Calvin cycle of photosynthetic organisms and the nonoxidative branch of the pentose phosphate pathway (PPP) (Figure 1.1 B) (Bernacchia *et al.*, 1995). Transketolase and transaldolase create a reversible link or bridge between the PP pathway and the glycolysis through their production of GA3P and F6P (Figure 1.1 B and C) (Sprenger et al., 1995, Wamelink et al., 2008).

The first observation of the transketolase enzyme was in rat's liver and spinach by Horecker and Smyrniotis in 1952 (Horecker et al., 1953)**.** Later on E. Racker and his team determined the dependence of transketolase enzyme from yeast on both the thiamin pyrophosphate and the magnesium ion (Racker et al., 1953). Josephson and Fraenkel were the first to isolate a transketolase mutant from *E. coli K12* in 1969. They discovered that their mutants lacking the transketolase activity were unable to grow on some carbon sources (e.g D-xylose) unless shikimic acid or an aromatic amino acid is added to growth medium (Josephson & Fraenkel, 1969). Several years were spent on studying transketolase in different organisms focused on *S. cerevisiae* (TKL-1), *E. coli* (tktA) and human erythrocytes and only in 1985 the first

sequence of transketolase gene from yeast *Hansenula polymorpha* was determined by Janowicz (Janowicz *et al.*, 1985).

Transketolase is found in all organisms investigated to date, from animals, yeast, plants to microorganisms and exists in the cytosol (Mathew, 2008). The first crystallography studies on the structure of transketolase was carried on yeast (Lindqvist et al., 1992) and *E. coli* (Littlechild et al., 1995) where both structures showed a striking similarity. The enzyme is a homodimer with two identical active sites. The two subunits of the enzyme consist of around 70 kDa and is build of three domains made of α/β type. The pyrophosphate binding domain (PP), the pyrimidine binding domain (PYR) and C-terminal domain, where only the PP and PYR domains are responsible for the active site formation as well as cofactor binding (the C-terminal domain has no active site function).

 The X-ray structure of transketolase domains from *E. coli* (E.C.2.2.1.1) showed that the PP domain is composed of five parallel strands of β-sheets and helices, the PYR domain made of six parallel strands of β-sheets cover of helices from both sides, the C-terminal domain consists of mixed β-sheets where four are parallel and one is antiparallel strand (Figure 1.2a and 1.2b) (Aucamp *et al.*, 2008, Kochetov, 2001).
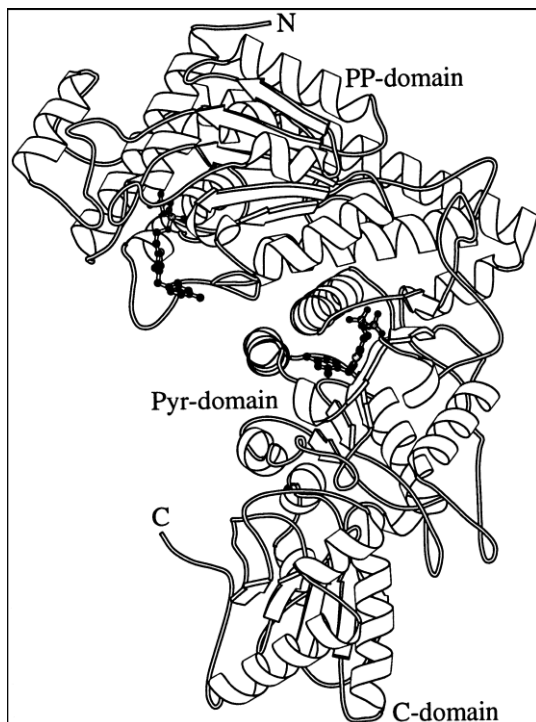
Figure 1.2a The E. coli transketolase subunit and the interaction of PYR and PP-domains with ThDP molecules. *The ThDP molecules are shown as ball and stick* (Schneider & Lindqvist, 1998).

Figure 1.2b The transketolase dimmer of E. coli (Schneider & Lindqvist, 1998).

The activity of transketolase depends on the thiamine pyrophosphate (TPP) and the divalent metal ion (e.g Mg^{2+}) as cofactors. The two ThDP molecules bind in the cleft located between the two subunits forming the active holoenzyme (Figure 1.2a). Thus the homodimer of the enzyme contains two of each cofactor (Racker et al., 1953). Other metal ions than Mg^{2+} can

bind, as Ca^{2+}, Mn^{2+} and Co^{2+} (Heinrich *et al.*, 1972). Kinetics studies on baker's yeast holo-transketolase showed it to be more stable with Ca^{2+} as a cofactor (Esakova *et al.*, 2005).

The transketolase enzyme is responsible for the transfer of a two carbon dihydroxyethyl groups from a ketose donor to an aldose acceptor. Thus the cycle of transketolase reaction consist of two parts. In the first half of the reaction, the donor substrate is being cleaved forming an aldose plus a covalently bound intermediate (α-carbanion). In the second half of the reaction the α-carbanion reacts with aldose (the acceptor) substrate. Hence, the two carbons from the first half reaction are transferred to acceptor substrate and ketose is formed by two carbon atoms added to its carbon backbone (Figure 1.3). Most transketolase enzymes able to utilize substrates of variable back bone length (C3-C7), this makes the enzyme important in synthetic organic chemistry (Bobest, 2004, Fiedler *et al.*, 2001, Wikner *et al.*, 1997).
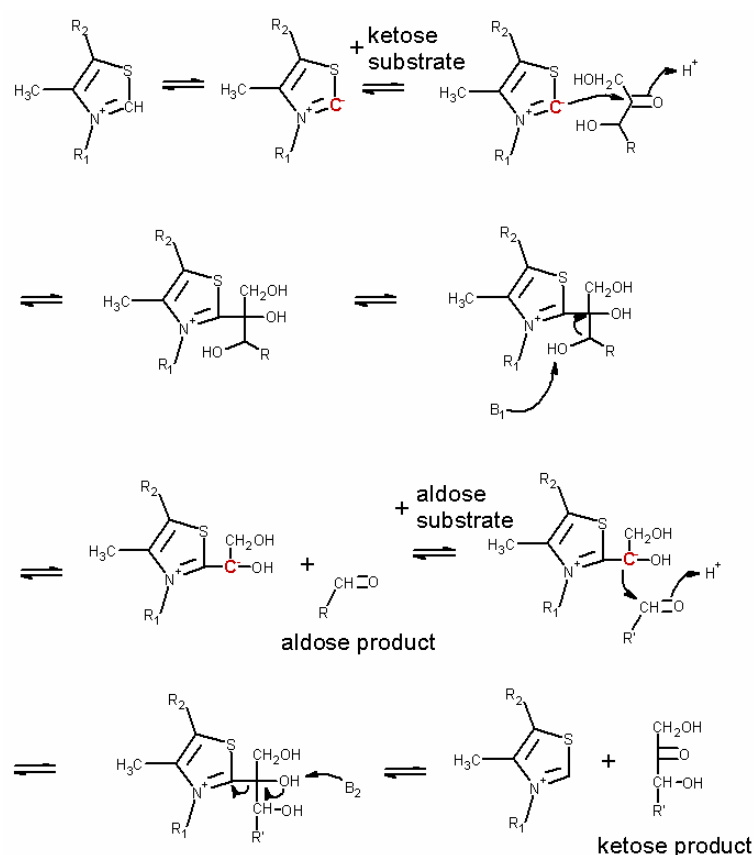
Figure 1.3 Trasnketolase reaction mechanisms (Bobest, 2004)***.*** *The anion at C2 position of ThDp is formed.The nuclephilic attach of the anion of the substrate forms a covalent adduct (a products formed after addition reaction). The product is released by cleavage of C-C bond induced by proton abstraction. The carbonations highlighted in bold, B1 and B2 indicate basic groups in active site, R1 and R2 indicate the pyrimidine and pyrophosphate groups of ThDP, R and R' indicates the remaining chains of the ketose and aldose substrates.*

Several reactions are carried by transketolase (Eidels & Osborn, 1971):

ribose-5-P + xylose-5-P $\rightarrow$ sedoheptulose-7-P + glyceraldehydes-3-P [1]

ribose-5-P + fructose-6-P $\rightarrow$ sedoheptulose-7-P + erythrose-4-P [2]

fructose-6-P + glyceraldehydes-3-P $\rightarrow$ xylose-5-P + erythrose-4-P [3]

Transketolase reactions listed above are reversible. It becomes irreversible if β - hydroxypyruvate is used to be a donor; the product is then carbon dioxide (Gyamerah, 1997).

Transketolase is an essential enzyme in the PPP, the products of its reaction serve as precursors for the synthesis of different biological molecules that function in cellular life (see chapter 1.1.2.1). Missing the activity of the transketolase due to mutations or inhibition would not lead to the synthesis of *ribose 5-phosphate* (for example), which is the first precursor of nucleic acid biosynthesis during cell division.

Several studies are carried out on mutants of this enzyme in yeast and prokaryotes as *E. coli*, *Corynebacterium ammoniagenes* and *Salmonella typhimurium*. Almost all showed that the deficiency in transketolase activity affected directly the biological processes in the cell. *E. coli* mutants lost the ability to utilize some carbon sources and failed to grow in medium lacking aromatic supplement (Josephson & Fraenkel, 1974). Experiments on modified transketolase in *C. ammoniagenes* concluded that the additional copy of transketolase decreased the inosine yield while deficiency of transketolase gave a higher yield of 5'-xanthylic acid (Kamada et al., 2001).

Transketolase exists in some organisms in several copies, yeast have been shown to have transketolase like-1 (TKL-1) and transketolase-like-2 (TLK-2) genes. *E. coli* contains transketolase A (*tktA*) which is expressed during exponential growth and transketolase B (*tktB*) expressed during stationary phase and showed to be responsible for 30% of the transketolase activity in the organism (Bobest, 2004) plants as *Craterostigma plantagineum* has three transketolase genes. In *A. vinelandii* there are three copies of the transketolase gene *tktA1*, *tktA2* and *tktA3* located in the chromosome.

1.1.3 Life cycle with focus on cyst formation

The life cycle of *A. vinelandii* was studied using electron microscopy (Figure 1.4) where three forms of the bacterium were observed, the vegetative cells, cyst and filterable cells. Lopez and Vela later discovered the filterable cells in 1981 (not shown in Figure 1.4). The filterable cells represent a normal stage of bacterias life cycle and are 0.3 μm in diameter, polymorphic rod-shaped cells. These cell of *Azotobacter* are able to pass through 0.45 μm membrane filter and to generate normal *Azotobacter* cells when grown in dialysed soil medium (Ballesteros *et al.*, 1986a). The filterable cells were observed in old cultures as their appearance occurs under unfavorable growth condition (Ballesteros *et al.*, 1986b). Thus, the resting cells of *A. vinelandii* have two forms, the filterable cellular stage and the cyst forming stage. Vegetative cells and the encystments is illustrated by Figure 1.4.
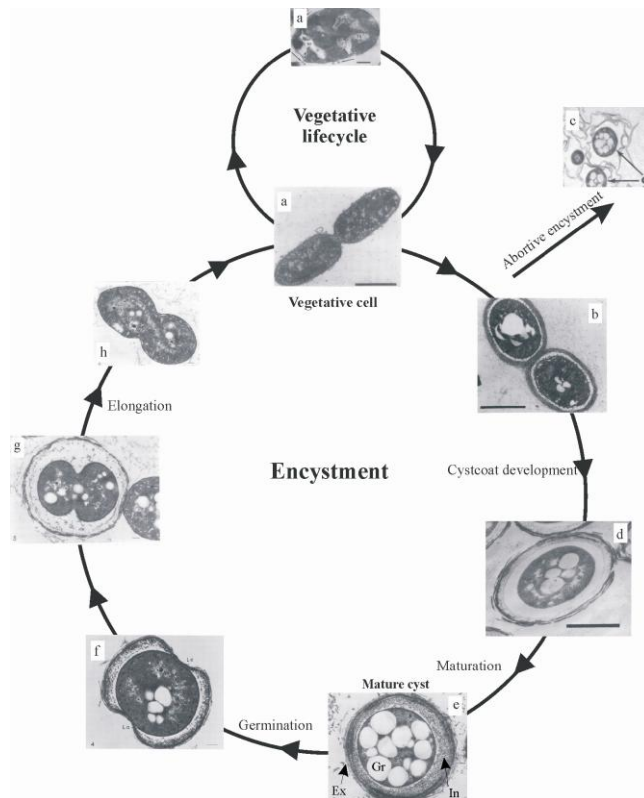
Figure 1.4 The life cycle of A. vinelandii (Steigedal, 2006) *a) vegetative cell, b) starting of encystment and accumulation of PHB granules c) aborted encystment d) exine development e) mature cyst, f) cyst germination g) initiation of cell division h) elongation and vegetative cell development. "In" indicates intine," Ex" exine, "Gr" PHB-granules. The life cycle is described in the text.*

The vegetative cells of this organism are mobile using flagella. Under cell division their shape become "peanut like" (Sadoff, 1975). During encystment, cells lose their motility due to disappearance of flagella (Hitchins & Sadoff, 1973) and become metabolically dormant. A single large rod-shaped vegetative cell gives rise to spherical cells as their walls become thicker over time (Wyss *et al.*, 1961). Cells stop their DNA synthesis and lower their nitrogen fixation process (Hitchins & Sadoff, 1973). The central body becomes embedded with a double layered coat or capsule, the layers are denoted the exine and intine. In the central body PHB accumulates (Socolofsky & Wyss, 1961).The cyst reaches its mature stage after 3 to 5 days, the cysts are half the size of the vegetative cell and its axes are 1.5 µm and 0.2 µm (Sadoff, 1975).

The intine (inner coat) and the exine (outer coat) are different in their structure, protein content, lipid distribution and mainly in alginate composition and quantity.
The exine consists of multiple layers and appears to be harder and stiffer because of its enrichment with alginate G-blocks (see alginate structure and properties section 1.2.1) (Lin & Sadoff, 1969).

The importance of cyst formation and cell dormancy in *Azotobacter* lies in their ability to protect the cell from its environment under unfavorable conditions and to germinate under suitable growth conditions. The cyst is more resistant than the parental vegetative cell to several environmental factors and harmful agents, such as ultraviolet irradiation, gamma radiation and desiccation. Desiccation is used as a tool to distinguish between cyst and the

vegetative cell in a mixture of *Azotobacter* cells (Socolofsky & Wyss, 1962). Neither cyst nor vegetative cells are resistant to heat (Sadoff, 1975) and each appears to have different chemical composition from the other, where cysts are higher in protein content and lower in both lipids and carbohydrates compared to vegetative cells (Lin & Sadoff, 1969).

Encystment in *Azotobacter* depends on two main factors to occur: alginate synthesis and the conversion in metabolism from carbohydrates to lipids. In some strains of *Azotobacter* cyst formation can be induced by adding carbon courses as hydroxybutyrate (Sadoff, 1975, Stevenson & Socolofsky, 1966). Both glucose and sucrose inhibits the formation of cyst (Sadoff, 1975).

Under favorable environmental conditions, the dormant cells germinate to become an active vegetative cell again. The volume of the central body of the cyst increases occupying the intine volume and the cells recover their motility prior to postgermination division (Sadoff, 1975, Loperfido & Sadoff, 1973).
The outgrowth of cyst from its outer cover is signaled by both DNA synthesis and nitrogen fixation (Loperfido & Sadoff, 1973). Under laboratory conditions *A. vinelandii* are able to germinate at 30°C in N-free Burk's medium with addition of appropriated carbon sources (Sadoff, 1975).

1.2   Alginate

1.2.1 Alginate structure and properties

Alginate originally was extracted from the brown algae in 1883 by the British chemist Stanford (Sabra & Zeng, 2009). In the1960's alginate synthesis by bacteria was discovered in *Pseudomonas aeruginosa* by Linker and Jones (Linker & Jones, 1964) and in *A. vinelandii* by Gorin and Spencer (Gorin & Spencer, 1966). Later on alginate biosynthesis was investigated in several *Pseudomonades* and *Azotobacter* species (Clementi, 1997).

The polysaccharide produced by these bacteria is an unbranched (1→4) copolymer composed of β-D-mannuronic acid (M) and its C-5 epimer α-L-guluronic acid (G) (Figure 1.5). Alginate is made of nonrepeating units and can be described as a mixture of three different block structures, G-blocks containing consecutive guluronate residues (G-residues), M-blocks containing consecutive mannuronate residues (M-residues) or MG blocks having alternatively M and G residues. The relative ratio and distribution of G and M makes the difference between alginates from different strains (Valla, 1998). All alginate producing organisms start making mannuronan, which become epimerized with C-5 epimerases to give rise to G residues (Valla *et al.*, 2001).
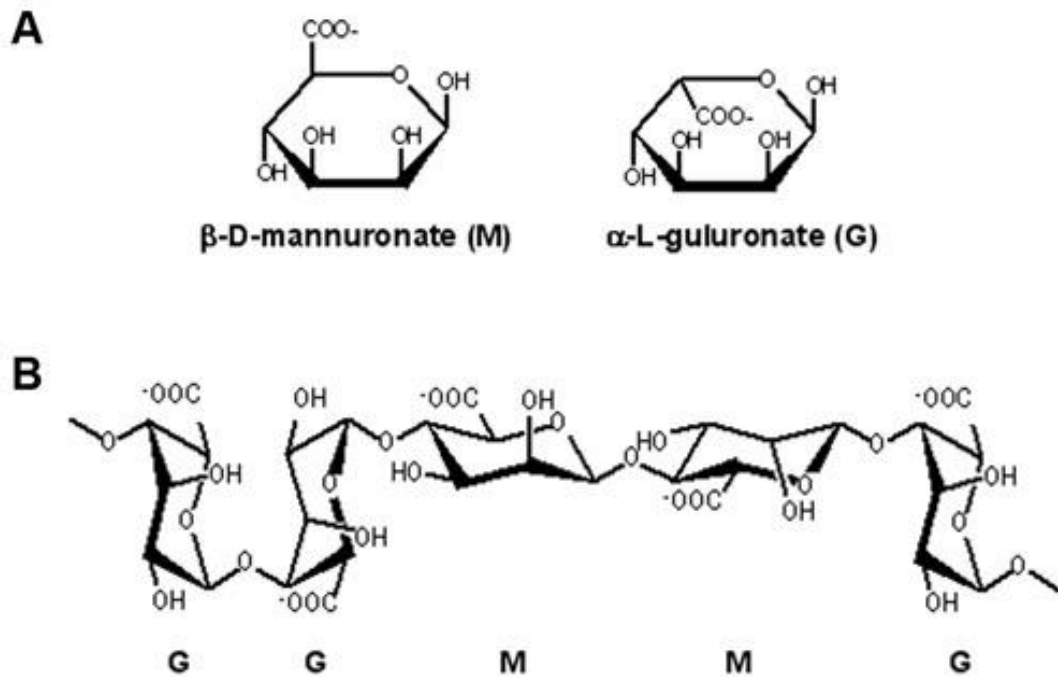
Figure 1.5 Alginate building blocks and its structure (Valla, 1998) *A) The two building block of alginate, β-D-mannuronic acid and α-L-guluronic acid. B) The structure the three block type; GG, MM and MG.*

Alginate is a negatively charged polymer and its two most important properties applied in industrial uses (industrial applications of alginate described in section 1.2.3) are the ability to form gels and viscous solutions. The properties of alginate are directly dependent on the structure and composition of the polymer (Smidsrød & Draget, 1996, Ertesvåg *et al.*, 1999).

The gelling property of alginate depends on blocks of L- guluronic acid (G) residues due to their responsibility for ion specific binding. The G-blocks are able to form strong gels in the presence of Ca^{2+} as divalent metal or H^{+} (Brownlee *et al.*, 2009, Stokke *et al.*, 1991). Gels can be formed under two physical conditions; the first is when M-residues responsible for gel formation in an acidic environment and the second is the selectivity of cations binding (Steigedal *et al.*, 2008, Donati & Paoletti, 2009)**.** The divalent cation (e.g Ca^{2+}) interacts with two adjacent G residues as well as with G-residues from opposing chains. This forms a junction zone, which serve as the physical interactions that connect the alginate chains together and lead to formation of hydrogel. The gel network is described as "The egg-box model" Figure 1.6 (Donati & Paoletti, 2009). Thus from this model the strength of the gel depends on the high content, distribution and the length of the G-blocks as well as the concentration of divalent cation in solution (Donati & Paoletti, 2009).
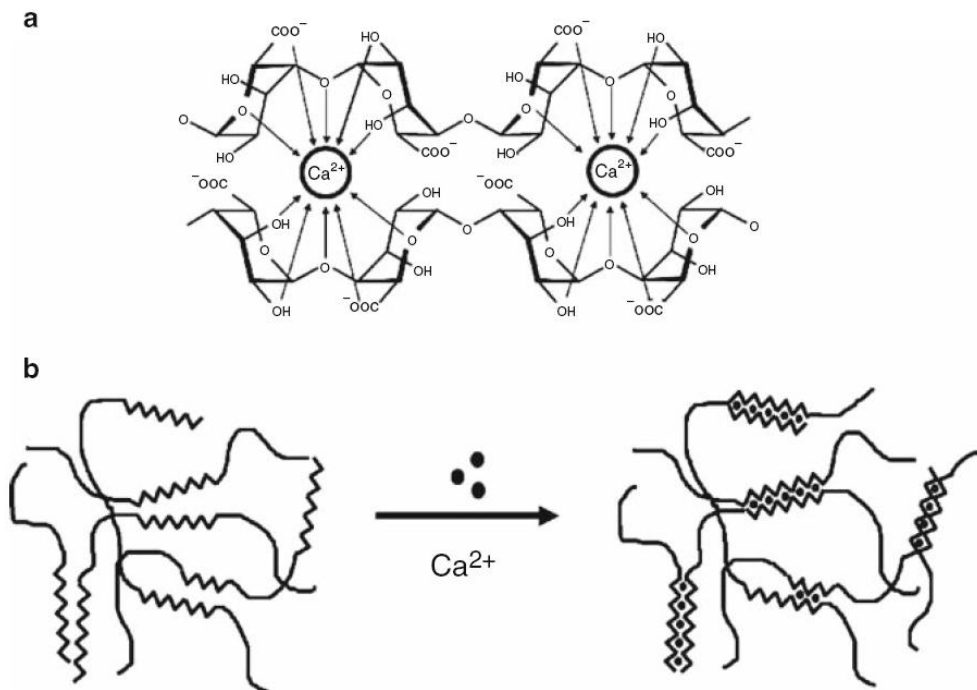
Figure 1.6 " The egg-box model" shows the binding of divalent cations to alginate G-blocks (Donati & Paoletti, 2009) *a)The positively charged divalent ions and the negatively charged region of alginates interaction to form junction of G-residues b) Interchain junction formation due to the addition of divalent cation (e.g Ca^{2+}) to alginate solution that leads to gel formation by G-blocks.*

The polymer length determines the viscosity of the alginate (Ertesvåg & Valla, 1998), as well as its molecular weight and composition. The composition of alginate blocks increase its flexibility ranking from GG>MM>MG (Donati & Paoletti, 2009).

Alginate produced by *Pseudomonades* lack the G blocks in contrast to found alginate from *Azotobacter* and algal alginate (Russell & Gacesa, 1988). While *Azotobacter* appears to have the same alginate composition as those of algal by containing M, G and MG-blocks and their G content can vary from 5-80% (Steigedal, 2006), there still a main difference between bacterial and algal alginate in the *O*-acetyl groups carried at O-2 and/or O-3 on the D-mannurosyl residues of the bacterial polymer. The presence of the acetyl groups on the bacterial alginate makes its polyelectrolyte behaviour different from the algal one, as the acetylation enhances alginate's capacity to bind water, which plays an important role under dehydration conditions (Steigedal, 2006, Rehm & Valla, 1997). The acetylation is also believed to protect the mannuronic acid from epimerization to guluronic acid as well as protecting the polymer from degradation by alginate lyase (Steigedal, 2006). In addition, bacterial alginate appears to have higher molecular weight than algal alginate. The degree of acetylation, molecular weight and the enzymes that controls these parameters are all directly involved in changing physical properties of alginate (Galindo *et al.*, 2007, Rehm, 2010).

10

1.2.2 Alginate biosynthesis

1.2.2.1 Genes of alginate biosynthesis and regulation

Alginate biosynthesis genes are arranged in a similar manner in both *Pseudomonades* and *Azotobacter* and they are all located on the chromosome. The two bacterial genera have some differences concerning their transcription, physical arrangement of genes and epimerization (Muhammadi & Ahmed, 2007). In *Pseudomonades* the genes cluster consists of 12 genes from *algD* to *algA*, all except *algC* are under control of a promotor localized upstream of *algD*. *algC* transcribed from its own promoter (Figure 1.7) (Remminghorst & Rehm, 2006).
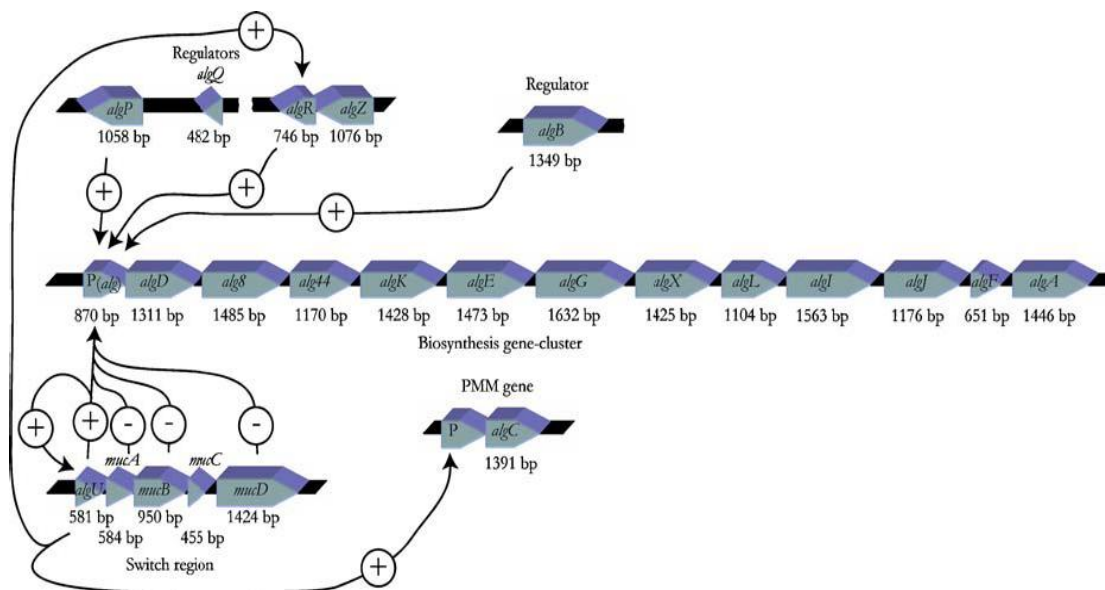
Figure 1.7 Genes of alginate biosynthesis and their regulation in P.aeruginosa (Remminghorst & Rehm, 2006). *Alginate genes are shown as arrows and are arranged in the direction of transcription,"+" for positive regulation, "-" for negative regulation.*

The biosynthetic gene cluster in *Azotobacter* described in Figure 1.8 is as well composed of 12 genes from *algD-8-44-K-J-G-X-L-I-V-F-A* with exception of *algC*. The genes are arranged in three operons where *algD* is transcribed alone under control of three promoters *algDp1*, *algDp2* and *AlgDp3*, *alg8* is under control of *alg8p* promoter located upstream to *alg8* gene and upstream to *algG* there is a putative sigma 70 promoter. *algC* gene has is transcribed from two own promoters (Campos *et al.*, 1996, Galindo et al., 2007).

Pseudomonades and *Azotobacter* contain the periplasmic AlgG protein for epimerization. In addition, *Azotobacter* is able to modify alginate by a family of extracellular Ca^{2+} dependent C-5 epimerases, which are not located within the cluster of alginate biosynthesis gene (Remminghorst & Rehm, 2006, Ertesvåg et al., 1999). The *algE* (later designed as *algE2*) was the first alginate gene cloned from *A. vinelandii* (Ertesvåg *et al.*, 1994). To date seven epimerase (*algE1-algE7*) genes belonging to the same family are identified in *A. vinelandii* (Bjerkan et al., 2004, Ertesvåg et al., 1999).
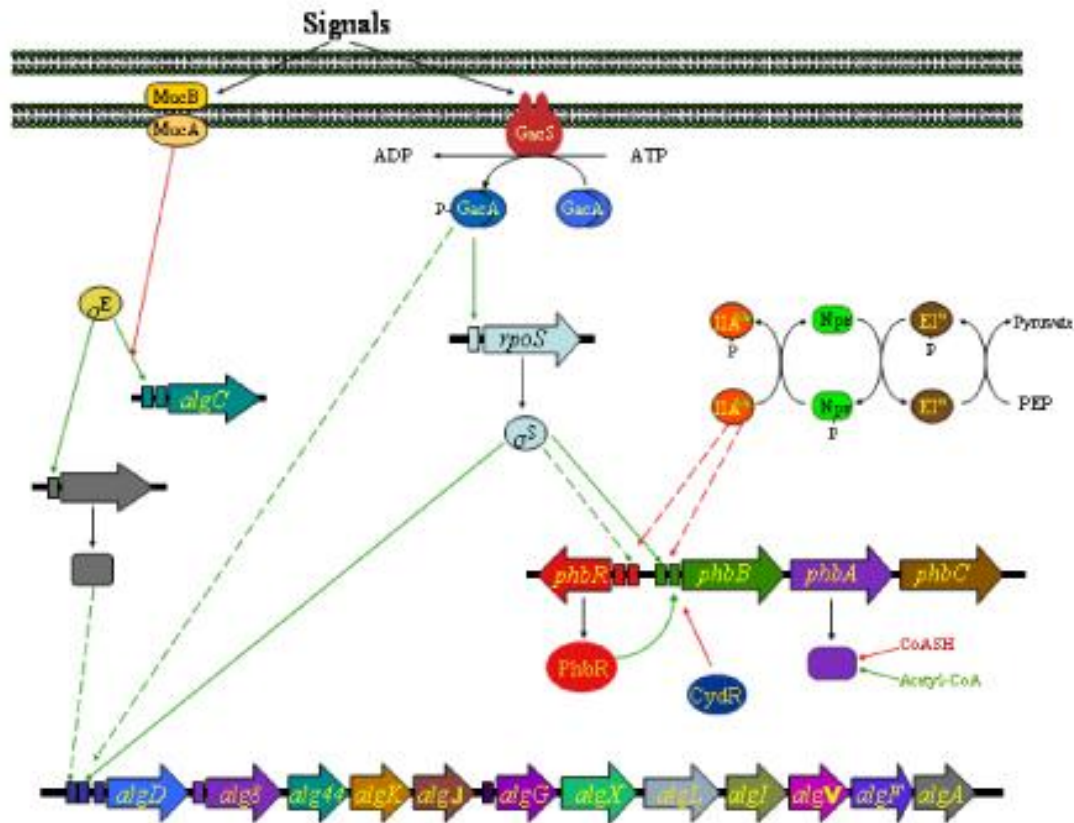
Figure 1.8 A. vinelandii alginate and PHB biosynthetic genes and their regulation (Galindo et al., 2007). G*enes in arrows are arranged in the site of transcription, rectangles indicate promoters, lines in green for positive regulation and lines in red for negative regulation. The picture was modified from the original in the reference, algJ and algV were added.*

The genes of alginate biosynthesis were first studied in *P. aeruginosa* due to its medical relevance as human pathogen, and can be divided into three clusters, **biosynthetic**, **genetic switching** and **regulatory genes** (Muhammadi & Ahmed, 2007). At least 13 genes are involved in the biosynthesis of alginate and 17 are regulatory genes (Steigedal, 2006).

From Figure 1.7, the regulatory mechanism of *P. aeruginosa* shows that *algT* (*algU*) is negatively regulated by products of *mucA*, *mucB*, and *muc D* genes while the role of *mucC* remains unclear (Hay et al., 2010). Therefore mutational inactivation of *mucA* and *mucB* genes lead to full activity of *algU* and over-production of alginate while inactivation of *algU* results in no synthesis of alginate. These genes also appear as genotypic switching genes controlling the conversation of *P. aeruginosa* from its non-mucoid to mucoid form (Muhammadi & Ahmed, 2007). In *A. vinelandii* the expression of the alginate biosynthetic genes also is under control of *algUmucABCD* gene cluster. *algU* code for alternative sigma E factor (σ^E), which is needed for transcription of the first promoter of *algC* (*algCp1*) and the second promoter of *algD* (*algDp2*) (Galindo et al., 2007).
In *P. aeruginosa* in responce to envelope stress the sigma factor is activated by regulated intermembrane proteolysis cascade (RIP). The key proteins of the alg/muc RIP are AlgU and the anti-sigma factor MucA. The degradation of MucA by the RIP cascade involves several proteases. In the responce to envelope stress MucA is cleaved by algW (periplasmic protease) as a result algU is released due to subsequental cleavage on its cytosolic side by YaeL

(intermembrane protease). The PDZ activating domain of algW is activated by a second protease protein, MucE (an outermembrane located protein). The third protease protein is MucD which appears to be involved in the degradation of misfolded proteins in the periplasm, where the misfolded proteins in the periplasmi activated algW (Figure 1.9) (Hay et al., 2010).
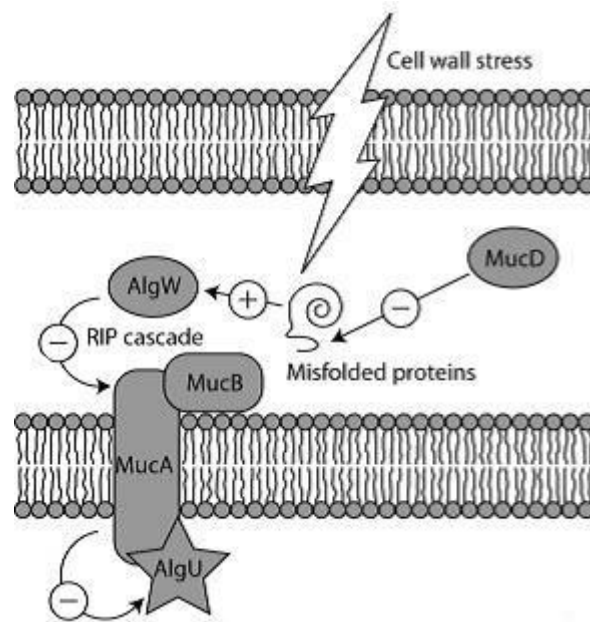
Figure 1.9 Overview of regulatory proteolytic cascade in P. aeruginosa (Hay et al., 2010).

In *P. aeruginosa* the regulatory genes *algR, algB, algQ algZ* and *algP* encode some proteins involved in the positive regulation of *algD* promoter (Figure 1.7) (Rehm & Valla, 1997). *alg44* contains in its N-terminus a c-di-GMP binding PilZ domain recently shown to be involved in the regulation of alginate biosynthesis in *P. aeruginosa.* The diguanylate synthesis are involved the MucR protein (Hay et al., 2010).

1.2.2.2 Bacterial alginate biosynthesis

In *Azotobacter* alginate biosynthesis occurs in active vegetative cells or as a part of cyst coat. The cyst formation play an important role in protecting the dormant bacterial cells from the environmental conditions (The cyst formation is described in section 1.1.3) (Clementi, 1997). Alginate synthesized by the active vegetative cells has other roles depending on different environmental conditions (Sabra & Zeng, 2009), one of which is the responsibility for adhesion of bacterial cells to surface (Read & Costerton, 1987).
P. aeruginosa is a human pathogen in lungs of patients with cystic fibrosis (CF). The bacteria adhere to epithelial tissues, colonize and invade tissues due to its ability to secret toxin and extracellular protease that enables it to destroy the host cell and its immune defence (Russell & Gacesa, 1988, Davies, 2002).

The first metabolic pathway of the alginate biosynthesis was determined in brown algae (Lin & Hassid, 1966) and firstly in *A. vinelandii* by Pindar and Bucke in 1975 (Pindar & Bucke, 1975).The general pathway from fructose to alginate seems to be identical in bacteria (*A. vinelandii* and *P. aeruginosa*) and algae (*Fucus gardneri*) (Clementi, 1997) where both

A. vinelandii and *P. aeruginosa* were showed to have the same biosynthetic mechanism (Rehm *et al.*, 1996).

The biosynthesis of alginate is divided into four major parts (Figure 1.10), (i) the synthesis of precursor fructose -6-phosphate in the cytoplasma, (ii) the cytoplasmatic membrane transfer and the polymerization, (iii) alginate modification in the periplasma and finally (iv) the export of alginate through the outer membrane (Remminghorst & Rehm, 2006).
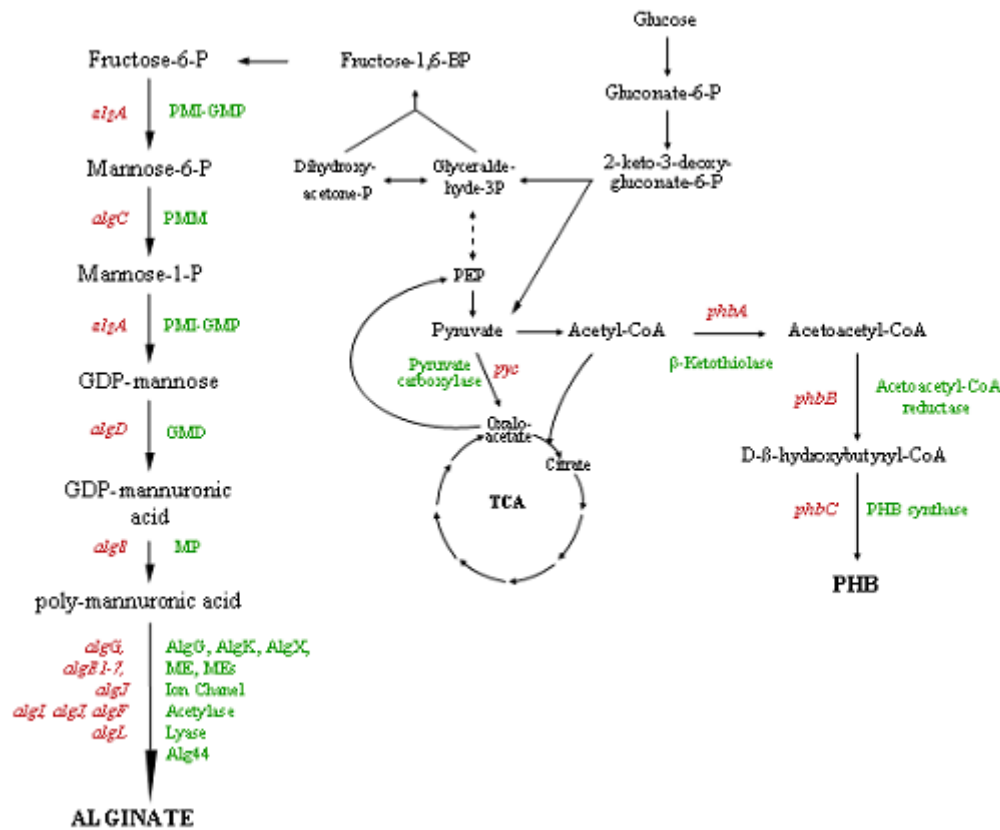
Figure 1.10 Biosynthesis of algiante and PHB in A. vinelandii (Galindo et al., 2007)**,** *genes for alginate and PHB biosynthesis indicated in red, the proteins responsible for the reactions indicated in green and described in details in the text.*

When the carbon source is glucose it utilized through the Entner-Doudoroff pathway and the synthesis of precursor which occurs in the cytoplasma starts from fructose-6-phosphate, which is converted to GDP-mannuronic acid through four enzymatic reactions. The first step is carried out by phosphomannose isomerase (PMI) encoded by *algA* gene where fructose-6-phosphates isomerized to mannose-6-phospahte. Mannose-6-phosphate is then converted to mannose-1-phosphate by the enzyme phosphomannonutase encoded by *algC*.
The GDP-mannose pyrophosphate activity of *algA* then converts mannose-1-phosphate to GDP-mannose. And finally GDP-mannose oxidized to GDP-mannuronic acid by guanosine-diphosphomannose dehydrogenase encoded by *algD* gene (Figure 1.10) (Remminghorst & Rehm, 2006, Muhammadi & Ahmed, 2007).

In the second step the GDP-mannuronic acid is transferred across the cytoplasmatic membrane. The polymerization process is carried out by alginate polymerase encoded by *Alg8* and *alg44* located in the periplasm (innen membrane) to produce its first product the polymannuronate (Remminghorst & Rehm, 2006, Muhammadi & Ahmed, 2007). In addition

14

Alg44 protein is proposed to be part of a periplasmic scaffold and to play role in bringing Alg8 in contact with AlgE (AlgJ in *A. vinelandii*) in the outer membrane (Figure 1.11) (Hay et al., 2010, Remminghorst & Rehm, 2006).

The polymannuronate then undergoes a number of modifications carried out by a number of periplasmic modification proteins such as; *algL* encode lyase for degradation thereby and cleaning the cell from excess production of alginate (Bakkevig *et al.*, 2005). For acetylation the proteins encoded by acetylase complex of *algI*, *algJ* and *algF* genes (*algI*, *algV* and *algF* in *A. vinelandii*) are needed and finally *algG* encode a protein responsible for epimerization (Figure 1.11) (Galindo et al., 2007, Rehm & Valla, 1997).

Finally the polymer passes its way out of the cell through the outer-membrane pore-forming protein AlgE encoded by *algE* gene (*algJ* in *A. vinelandii*) (Galindo et al., 2007, Rehm & Valla, 1997). Where the products encoded by *algX* and *algK* genes remain unclear but necessary for alginate production and believed to protect the alginate through the periplasmic space from modification proteins (Hay et al., 2010). In addition, AlgG is proposed to be needed in transport of the polymer through the periplasm (Figure 1.11) (Gimmestad *et al.*, 2003).
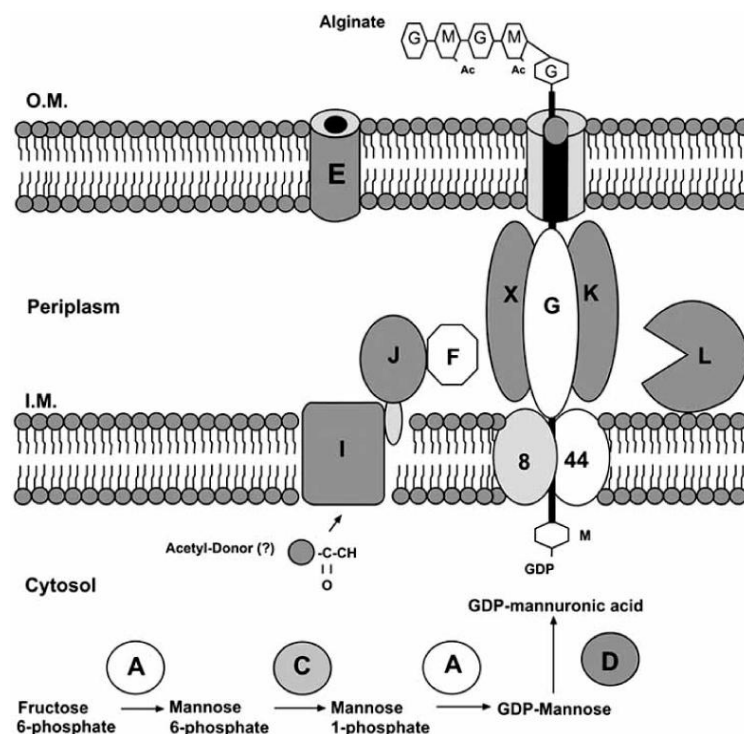
***Figure 1.11 The general model of alginate biosynthesis and its transport to the environment* (exemplified by *P. aeruginosa*)** (Muhammadi & Ahmed, 2007)*. The O.M. is oute-membrane, I.M. is inner membrane, A, C, D, 8, 44, G. etc are alginate biosynthesis genes (algA, algC and so on), AC indicates acetyl group.*

1.2.2.3 Factors affect alginate production in *A. vinelandii*

Oxygen concentration influences the molecular weight and the production of alginate in *A. vinelandii* (Sabra *et al.*, 1999, Pena *et al.*, 2000). In addition to alginate the bacteria produces PHB (see section 1.1.1), this production process appears to be a competing process (Steigedal, 2006). The limitation of oxygen increases PHB synthesis and lowers the production of alginate. The production of alginate can also be lowered by high oxygen level (Steigedal, 2006, Parente E., 1998). High production rate of alginate was determined by the addition of iron and/or molybdate where both are important in the nitrogen fixation process of the bacteria. Some studies showed that the alginate production was inhibited by the addition of ammonium (Steigedal, 2006).

1.2.3 Industrial and pharmaceutical applications of alginate

The importance of alginate in medical and industrial application comes from their properties as stabilizing, gelling and viscosifying agent as well as their ability to retain water (Rehm, 2010). The alginate used for commercial purposes are currently extracted from brown algae. Production of alginate by microorganisms by fermentation is possible but still expensive commercially (Draget & Cathrine, 2011, Sabra et al., 2001).
Traditionally alginate was used as thickening agent in both textile industry and food applications (Steigedal, 2006) in ice cream, jam, and sauces including mayonnaise (Brownlee *et al.*, 2009). The biomedical and pharmaceutical applications such as dental materials, dressing in surgery, and both dermal and epidermal wounds management and medical drugs have been increasing (Draget & Cathrine, 2011, Remminghorst & Rehm, 2006). Alginate is used in oral tablets due to their property of forming gels in acid environment (gastric). This gel formation around the tablet is able to shrink at acidic environment (stomach) protecting the tablet's compounds and acting as buffering agent when tables are made using Ca or Na-alginate. The alginate surrounding the oral tables erode in alkaline pH, this is used to deliver drugs into the intestine of pH > 6.7. The other feature of oral tablets trapped by alginate is that the polymer is adhesive and thus able to attach to intestine mucosa for a long period of time (Draget & Cathrine, 2011, Qurrat ul *et al.*, 2003).
The most advanced biotechnological and medical application is a technique of encapsulation of living cells with Ca-alginate shell (gel-spheres) and their used in transplantation (Donati & Paoletti, 2009). The gel around the immobilized cells works as a selective immune barrier protecting the transplanted cells from the immune system of the host (Draget & Cathrine, 2011). The encapsulation system have been applied in diabetes mellitus, liver failure, hemophilia B, colon and brain cancer as well as cells for testosterone supplement (Steigedal, 2006).

1.3 Expression system in *E. coli* with focus on *Pm-promoter*

The expression of recombinant proteins in bacteria is one of oldest technique in recombinant DNA technology, where the recombinant gene of interest is cloned under a control of inducible promoter in an expression vector and transformed into bacteria host (Brautaset T., 2009).

E. coli is one of the mostly used hosts in expression systems for production of recombinant proteins in bacteria since its well understood and characterized both physically and genetically. The bacterium has a rapid doubling time (approximately 30 min) and a small genome size (4,639,221 bp). *E. coli* cells are easily lysed for the harvesting of the protein produced within it (Lale, 2009, Reece, 2004).

Regulation of promoter activity can be obtained by changing and optimizing environmental conditions such as; temperature, pH and the ligands. These factors affect the transcriptional regulators that may enhance or repress or both positively and negatively regulate the transcription from a specific promoter (Brautaset T., 2009). The promoter can be regulated by repressor or activator proteins.

The inducible promoter complex stimulates the transcription through binding to a specific site in the DNA upstream the promoter. This binding enables the interaction with polymerases as a result the transcription becomes facilitated. This activation happens when the inducer binds to repressor, preventing its binding to operator and allowing the transcription by RNA polymerases (RNAP). While in the absence of the inducer, the repressor binds to operator sequence hiding the RNAP binding site thus inhibiting the initiation of transcription (Brautaset T., 2009).

One of the promoter systems used in high level expression of recombinant protein in gram negative bacteria species is the *Pm-xylS*.

The *Pseudomonades putida* TOL (toluene degradation) plasmid pWW0 and encodes two operons, the upper-operon encodes enzymes involved in the conversion of substrates to benzoate or alkylbenzoate. And the meta-operon that drives the enzymatic pathway for conversion of aromatic compounds (3-methylbenzoat) into an intermediate in the Krebs cycle, this pathway is called "meta-cleavage pathway" or lower pathway (Winther-Larsen *et al.*, 2000).

The transcription of meta-operon from *Pm* promoters which is 117 kb (Lale, 2009) is positively regulated by *XylS* gene. The *XylS* belongs to AraC/XylS transcription factor family (321 aa long) and is transcribed from two individually regulatory promoters the *Ps1* and *Ps2*. The positive regulation of *Pm* promoter by *XylS* genes carried out the formation of an active complex in the presence of inducer such as benzoate or it derivatives (e.g *m*-toluate) (Figure 1.12). XylS protein recognizes the *Pm* promoter in two regions placed -70 and -35 with respect to the initiation site of transcription (Marques *et al.*, 1999, Winther-Larsen et al., 2000). The expression of *Pm* promoter can as well enhanced without the addition of inducer by excess amounts of XylS protein (Lale, 2009).
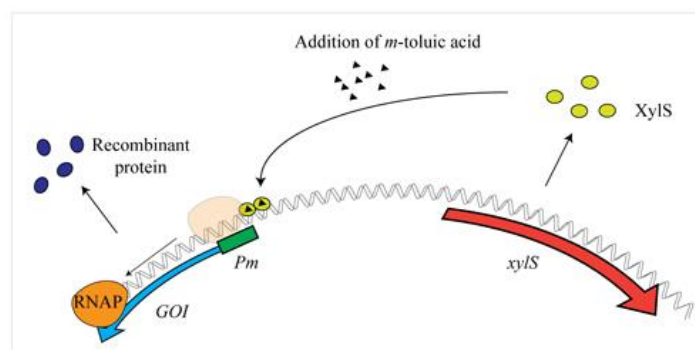
Figure 1.12 The positive regulation of Pm-promoter by XylS gene using m-toluic acid as effecter (Vectronbiosolutions, 2010).

The *Pm* –promoter is well studied and a number of works published for the improvement of its function through mutations or other induced techniques. The *Pm* –promoter used in this work is both the wild type cloned from the *P. putida* TOL plasmid and a G5 mutant for the meta-cleavage operon. The *PmG5* mutant promoter is a derivative of *Pm* wild type promoter. The mutant *PmG5* in the absence of inducer lowers the background expression comparing to the wild-type *Pm*-promoter (Bakkevig et al., 2005).

1.4 Aim of the study

Alginate is an important industrial polysaccharide produced by *Pseudomonades* and *Azotobacter*. To be able to increase the yield of alginate production to meet the industrial uses, the genes involved in its biosynthesis have to be studied. As the genome of *A. vinelandii* is already sequenced it was constructed a mutant library, where random genes have been knock out by a transposon. The transposon mutant library was constructed using the TnCAM140 transposon in *A. vinelandii* ATCC12518Tc strain. In this strain a tetracycline resistance gene has been inserted in a gene encoding a transpoase (Ertesvåg, unpublished). The transposon contains a spectinomycin resistance gene and the reporter gene *β*-glucuronidase (*uidA*) (Wilson *et al.*, 1995).

A. vinelandii encodes three transketolase genes, which was confirmed by genome sequencing. The study carried out by Vie, A. K. in 2008 on transposon mutant library resulted in the identification of a transposon mutant of *tktA* (*tktA3*) which is believed to be involved in the reduction of alginate production up to 50% and inhibition of cyst formation in *A. vinelandii*. To determine whether the observed phenotype is due to transketolase mutation and not related to the other gene found in the same operon, it was suggested to construct an in frame deletion in order to study this phenotype without affecting the expression of the other genes in the operon.

The aim of the present work was proposed to study the influence of the three mutant transketolase (*tktA1*⁻, *tktA2*⁻ and *tktA3*⁻) genes on alginate biosynthesis and cyst formation in *A. vinelandii*.

In order to study the effect of transketolase genes *in vivo* on the biosynthetic process of alginate synthesis, the in frame deletion in three transketolase genes was constructed using oligonucleotide primers. The flanking sequences were cloned into suicide vectors. Through a homologous gene replacement, the three transketolase genes were disrupted in *A. vinelandii*. The complementation analyses with a wild-type transketolase have to be carried out, in order to make sure that the observed phenotype is due to lack of transketolase. The complementation study was planned to be carried using suicide transposon vector for complementation, which is able to insert the wild-type version of the gene into the chromosome of *A. vinelandii* carrying the mutant gene.

The second aim of the study was to analyse the transketolase activity of *A. vinelandii in vitro*, by cloning, characterizing and carrying an expression analysis in *E. coli*. The three transketolase wild-type genes were cloned under the control of *Pm*-promoter and overexpressed in *K-12 E. coli* strain.
The activity of the three genes was determined by enzyme assay and SDS-PAGE. The three transketolase genes were purified by FPLC and their concentration estimated.

a)

```
┌─────────────────────────────────────────────┐
│  Experimental study of three transketolase   │
│                 mutant genes                  │
└─────────────────────────────────────────────┘
```

```
┌──────────────────────────┐      ┌──────────────────────────┐
│ Construction of plasmids  │      │ Construction of transposon│
│  for gene inactivation    │      │ complementation vectors   │
└──────────────────────────┘      └──────────────────────────┘
```

| pMK5 | pMK9 | pMK12 | pMK3 | pMK7 | pMK10 |

```
┌──────────────────────────────────────┐
│ Gene inactivation in A. vinelandii    │
│ through homologous gene replacement   │
└──────────────────────────────────────┘
```

```
┌───────────────────────────────────────────────────────┐
│               Complementation study                    │
└───────────────────────────────────────────────────────┘
```

b)

```
┌───────────────────────────────────────────────────────┐
│ Cloning of three A.vinelandii wild type transketolase  │
│ genes (tktA1, tktA2 and tktA3) in E. coli              │
└───────────────────────────────────────────────────────┘
```

| pMK13 | pMK14 | pMK15 |

```
┌───────────────────────────────────────────────────────┐
│   Protien purification and  expression analysis        │
└───────────────────────────────────────────────────────┘
```

Figure 1.13 scheme of the experimental part of the work a) *study of three A. vinelandii mutant transketolase genes and their influence on alginate biosynthesis and cyst formation **b)** Cloning and expression analysis of three A. vinelandii wild type transketolase gene in E. coli.*

2 Methods and Materials

2.1 Bacteria Strains and plasmids

Plasmids and bacteria strains used in this work are listed in table 2.1. Detailed plasmid map is shown in appendix B.

Table 2.1 Bacteria strains and plasmids used in this work with their description and source.

Bacteria strain	Description	Source
A. vinelandii ATCC 12518	Wild-type	**American Type Culture Collection**
E. coli S 17.1 λ pir	R6K λ phage, lysogenic strain of *E. coli S17.1*	**de Lorenzo et al, 1993**
E. coli S17.1	*E. coli* 294 with integrated plasmid RP4-2 Tc::Mu-Km::Tn7.containing *tra* and *trfA* genes	**Simon et al., 1993**
RV308	(Lac)x74 galPO-308::IS2 rpsL	**Bethesda Research Laboratories** **Li, McClure at al 1997**
Plasmids		
pEM2	pLitmus-28 vector with *trfA* 254D cloned down stream of *PmG5*- promoter. Amp^R	**Mellemsæther E. 2006**
pHE272	Contains transposon –trp*, *bla, Pm*-promoter, *xylS*, *aadA*, oriT.	**Ertesvåg H.**
pHE263	Contains *Pm*-promoter, *xylS, algL*, M13-ori, oriV, *bla*	**Ertesvåg H.**
pAKV1	Contains *Av2301* cloned in lacZ	**Vie A.K. 2008**
pHE303	Derivative of pUC128 in which a 1.1 kb PCR fragment encoding *opptktA1* from *A. vinelandii* ATCC 12518 cloned in *lac*Z	**Ertesvåg H.**
pHE304	Derivative of pUC128 in which a 1.1 kb PCR	**Ertesvåg H.**

	fragments encoding *opptktA2* from *A. vinelandii* ATCC 12518 cloned in *lac*Z	
pHE216	Contains *algE3*, *Pm*-promoter, *xylS*, *bla*, oriV, *sacB*, *aadA*.	**Ertesvåg H.**
pMK2	Derivative of pEM2 in which a 2.1 kb NotI/NdeI DNA-fragments of *tktA1* from *A. vinelandii* ATCC 12518 was cloned down stream of the *PmG5*-promoter. Amp^R	**This work**
pMK3	Derivative of pHE272 in which a *tktA1* NotI/NdeI fragment from pMK2 was inserted down stream of the *PmG5* –promoter. Sp^R	**This work**
pMK4	Derivative of pHE304 in which a 1.3 kb SacI/PsiI fragment from pMK2 was cloned in *lac*Z. Amp^R	**This work**
pMK5	Derivative of pMK12 in which a 2.0 kb SalI/NotI fragment from pMK16 was cloned in *lac*Z. Amp^R, *sacB*	**This work**
pMK6	Derivative of pEM2 in which a 2.1 kb NotI/NdeI PCR DNA-fragment encoding *tktA2* from *A. vinelandii* ATCC 12518 was cloned down stream of the *PmG5*-promoter. Amp^R	**This work**
pMK7	Derivative of pHE272 in which a *tktA2* NotI/NdeI fragment from pMK6 was inserted down stream of the *PmG5* –promoter. Sp^R	**This work**
pMK8	Derivative of pHE304 in which a 1.3 kb SacI/PsiI fragment from pMK6 was cloned in *lac*Z. Amp^R	**This work**

pMK9	Derivative of pMK12 in which a 2.0 kb SalI/NotI fragment from pMK17 was cloned in *lac*Z. Amp^R, *sacB*	**This work**
pMK10	Derivative of pHE272 in which a *tktA3* NotI/NdeI fragmenst from pAKV1 was inserted down stream of *PmG5* –promoter. Sp^R	**This work**
pMK11	Derivative of pAKV1 in which a 1.7 kb NotI/NdeI PCR fragment encoding *opptktA3* from *A. vinelandii* ATCC 12518 was cloned down stream of the *PmG5*-promoter. Amp^R	**This work**
pMK12	Derivative of pHE216 in which *opptktA3* SacI/PspOMI fragment from pMK11 was cloned down stream of the *Pm-promoter*. Amp^R, *sacB*	**This work**
pMK13	Derivative of pHE263 in which an insert of *tktA3* NotI/NdeI from pMK10 was cloned down stream of the *Pm-promoter*. Amp^R	**This work**
pMK14	Derivative of pHE263 in which an insert of *tktA1* NotI/NdeI from pMK2 was cloned down stream of the *Pm-promoter*. Amp^R	**This work**
pMK15	Derivative of pHE263 in which an insert of *tktA2* NotI/NdeI from pMK6 was cloned down stream of the *Pm*-promoter. Amp^R	**This work**
pMK16	Derivative of pMK4 in which a SacI DNA-fragments was removed. AmpR	**This work**

| pMK17 | Derivative of pMK8 in which a SacI DNA-fragments was removed. AmpR | **This work** |

The three constructed deletion mutation in the transketolase genes (*tkt1⁻*, *tktA2⁻* and *tktA3⁻*) are described in table 2.2. Table 2.2 shows the size of wild-type transketolase genes in base pairs comparing to the deletion mutants as well as the amount of amino acids deleted and their location in each of the three genes.

Table 2.2 The three transketolase wild-type genes compared to the deletion mutants constructed

Gene	*Deletion in amino acids residues*	*Size in kilo base and amino acids*
tktA1 (WT)*		2.1 / 672 aa
tktA1 deletion mutant	189 to 672	1.4 / 189aa
tktA2 (WT)		2.1 / 672 aa
tktA2 deletion mutant	189 to 672	1.4 / 189 aa
tktA3 (WT)		2.0 / 665 aa
tktA3 deletion mutant	1 to 440	1.7 / 234 aa

* WT is wild-type

2.2 Experimental methods

2.2.1 Preparation of cells

Cells were taken with a sterilized tooth pin from a frozen cell culture or agar plate and transferred to vials with growth medium. *E. coli* cells were grown in LB medium at 37°C. *A. vinelandii* was grown on Burks medium (BM) or RA1 at 30°C.

2.2.2 Storage of cells /cell culture

2.2.2.1 Long time storage

Cells were allowed to grow until the stationary phase before they were transferred into cryo-tubes. 300 µl of autoclaved 60% glycerol per each 1 ml of culture was added. The cryo-tubes were stored at -80°C.

2.2.2.2 Short time storage for use in PCR and sequences reaction

Cells were allowed to grow until they reach a stationary phase, then 1 ml of culture transferred to 1.5 ml tube and was centrifuged for 10 min at 10.000 x g, the supernatant was discharged and the pellet was resuspended in 100 µl dH$_2$O. The cells cooked at 100°C for 10 min. and stored at -20°C.

2.2.3 DNA isolation

2.2.3.1 Isolation of plasmid DNA

The isolation of plasmids were carried out using *Wizard SV minipreps kit protocol*. The system can be used to separate plasmids from 1-10 ml culture. The system is mostly effective for plasmid from *E. coli* and works best with plasmids less than 20,000 bp in size (Promega, 2007).

Procedure

Cells were harvested by centrifugation for 5 min at 10.000 x g in a table centrifuge of 1.5 ml (high copy plasmid) and 10 ml (low copy plasmid) of overnight culture of *E. coli* cells. The supernatant was pourd off and the tube inverted on a paper towel to remove the excess media. The cells were lysed by addition of 250 µl of Cell Resuspension Solution and the cells were completely resuspended by pipetting. 250 µl of Cell Lysis Solution was added to bacteria lysate and mixed by inverting the tube 4 times, 10 µl of Alkaline Protease Solution was added to the bacteria lysate and mixed well. The lysate was incubated for 5 min. and 350 µl of Neutralization Solution was added to the bacteria lysate and mixed by inverting the tube 4 times. The bacteria lysate were centrifuged at max speed (14.000 x g) for 10 min at room temperature (Promega, 2007).

The purification of plasmid DNA from bacteria lysate use a spin column inserted into one 2 ml collection tube. The cleared lysate was transferred by decanting to the spin column. The supernatant centrifuged at max. speed for 1 min at room and the flowthrought was discarded

from the collection tube. The spin column was reinserted into the tube. 750 µl of Column wash Solution was added to spin column and was centrifuged at maximum speed for 1 min, the flowthrogh was discarded and the spin column was reinserted into the collection tube. The washing procedure was repeated using 250 µl of Column Wash Solution and the bacteria lysate was centrifuged for 2 min. The spin column was transferred into new 1.5 ml Eppendorf tube (microcentrifuge tube) and the plasmid DNA was eluted by adding 100 µl of Nuclease-Free Water to the spin column and the tube was centrifuged for 1 min. The microcentrifuged tube with purified plasmid DNA was stored at -20°C (Promega, 2007).

2.2.3.2 Isolation of genomic DNA from *A. vinelandii*

The genomic DNA from *A. vinelandii* was isolated following the PurElute^{TM} Bacterial Genomic Kit from EdgeBio with some modification on the buffers and solutes.

10 ml of stationary phase culture was centrifuged for 10 min at 5000 rpm, and the cells were washed by resuspention in 10 ml 0.9%NaCl, 10mM EDTA and were centrifuged again for 10 min at 5000 rpm.
The supernatant was discarded and the cells were resuspended in 400 µl Spheroblast buffer and were incubated at 37°C for 1 hour.
100 µl of 10% SDS and 100 µl of 5M NaCl were added and the solution was mixed and incubated at 65°C for 10 min.
100 µl of Advamax-pearls and 100 µl of 0.72 M $MgCl_2$ were added and mixed with the solution and the mixture was incubated at 30°C in shacking incubator for 1.5 hour.
The mixture was centrifuged at 13000 rpm at 4°C for 20 min, and the supernatant was transferred to a new sterile tube.
750 µl of isopropanol was added and was mixed by inverting the tube until the DNA precipitate. The mixture was centrifuged at 13 000 rpm for 25 min.
The pellet was washed with 300 µl 70% ethanol and was centrifuged at 13 000 rpm for 10 min. All the supernatant was pipetted out and the pellet was dried at room temperature for 1 hour.
25 µl of TE (pH 8.0) was added and the tube was placed in the shacking incubator at 37°C over night. More TE can be added if the pellet is not dissolved completely.

2.2.4 Amplification of DNA fragments using PCR

PCR (Polymerase Chain Reaction) is a method of amplification of specific DNA fragments; the method was invented by Kary Mullis in 1985 (McPherson & Møller, 2006).

PCR copies DNA in three separated stages using different temperatures .The reaction requires several components in addition to the DNA template, as two oligonucleotide primers 16-30 nucleotides long, buffer, dNTP and enzyme (e.g *Taq* DNA polymerase) (McPherson & Møller, 2006, Reece, 2004).

The stages of PCR reaction;

- Denaturation: The double-stranded DNA is separated (denatured) at 94°C. The temperature should not be too high in order to keep the enzyme (e.g. DNA polymerase) active.

- Anneling: The reaction is cooled to the anneling temperature at which the oligonucleotid primers hybridize to DNA template. The anneling depends on the length of the primers and typically in the range of 50-60°C.

- DNA synthesis or extension: the temperature is raised to 72°C and the thermostable DNA polymerase binds to the free 3-end of the primer and elongates the DNA strand in the 5'to 3' direction.

The ability of PCR to amplify a specific defined region even a very complex DNA as genomic DNA relay in the two oligonucleotide primers, which acts as a site for DNA synthesis. The primers are complementary to regions of known sequence on opposite DNA strands. The template and their free 3-end pointed toward the other primer, in a way that the region between the two primers will be amplified.
Primers are synthesized oligonucleotides ranges from 16-30 nt with a good specificity to a target DNA sequence. For a successful PCR reaction, the primers have to be designed carefully considering some factors and have to full fit certain criteria to be recognized as sensitive and high specific. The primers have to be designed avoiding repetitive sequences or stretches containing the same nucleotide as this can results in slipping the primer on the template. In designing the primers more than three G or Cs at the 3'end have to be avoided, as this can lead to mispriming at GC rich region as well as not containing sequences at 3'end that may lead to misparing with it self or other primer in the PCR reaction as this results in the formation of primer-dimer. The primers have to be specific and perfectly matched to the template in their 3'end to be able to distinguish, the target DNA sequence as this end of the primer that extended by DNA polymerase as ensures the specificity of annealing to the DNA target sequence of interest. The 5' end of the primer appears to be less important concerning the specificity of anneling the template. This helps the researchers to design primers with modifications at the 5'end as mutagenesis restriction sites that are needed in cloning procedures (McPherson & Møller, 2006).

Tables 2.2 and 2.3 contain a list over primers and their oligonucleotide sequences used in this work. The primers were designed using Clone Manager and ordered from Sigma-Aldrich. Two groups of primers are used; primers for amplification of transketolase genes *tktA1, tktA2,* and *opptktA3* from *A. vinelandii* ATCC12518 and primers for sequencing of PCR products.

Table 2.3 primers used for amplification of transketolase genes *tktA1*, *tktA2* and *tktA3*.

Primers	*Oligonucleotide sequences*
For amplification of tktA1 gene tktA112ATG and tktNotI	5' GATAACGCATATGACCCAGAACCTGAAACC 3' 5' ATAGCGGCCGCAACATCCAGTGCTCGTACAG 3'
For amplification of tktA2 gene tktA12ATG and tktA2Not1B	5' TATGCGGCCGCTCCGGATCGAGTATCAC 3' 5' TATGCGGCCGCTCCGGATCGAGTATCAC 3'
*For amplification of opptktA3 gen*e tktA3nedF and tktA3nedR	5' ATTAGATCTGCGCCACCTTCCTGGTGTTC 3' 5' ATTACTAGTTCGGTGTAGTCCAGCAATC 3'

Table 2.4 primers used for the sequencing of PCR products in the designated plasmids

Primers	Oligonucleotide sequences
For sequencing of tktA1 PCR product	
M 13 rev	5' AGCGGATAACAATTTCACACAGGA 3'
tktA1oppR	5' CGAAGCCGATGGTGGTCTTG 3'
tktA24767	5' CGAAGCCGATGGTGGTCTTG 3'
Pm forw	5' GACGACAACGGCATCTCCATC 3'
tktA12ATG	5' TTGCGTAGATCCGGTCGAG 3'
tktA1NotIR	Typed above
	5' ATAGCGGCCGCAACATCCAGTGCTCGTACAG 3'
For sequencing of tktA2 PCR product	
M 13 rev	Typed above
tktA1oppR	Typed above
tktA24767	5' GACGACAACGGCATCTCCATC 3'
Pm forw	Typed above
tktA12ATG	Typed above
tktA2Not1B	Typed above
For sequencing of opptktA3 PCR product	
M13 rev	Typed above
tktA24767	Typed above
tktA3 nedF	Typed above
tktA3 nedR	Typed above
For sequencing of tktA3⁻ mutant PCR products in A.v.ATCC12518:pMK12	
2301 NdeI	5' TAATAACATATGCCCAGCCGTCGTGAGCGA 3'
2301 NsiI	5' TTAATGCATGACAGGTCTCGGAGGTAG 3'

Not the primers alone have to be taken under consideration in the PCR reaction, but the buffers as magnesium, KCl and Tris-HCl as well (McPherson & Møller, 2006).

In this work Pwo SuperYield-kit was used and the sample was incubated for 5 min at 95°C and the thermocycling conditions were as follows;
30sec denaturing step 95°C, 30 sec annealing at 55°C and 2 min and 30 sec extension step at 72°C for fragments of 2.2 kb (the amplified fragments ranged from 1.1 kb to 2.2 kb, thus 1 min per 1 kb). This was repeated for 35 cycles.

Reaction mixture
5 µl Pwo SuperYield PCR buffer, 10x
1 µl dNTP, 10mM
1 µl fwd primer
1 µl rev primer

10 µl GC-rich resolution solution 5x
0.5 µl Pwo SuperYield DNA polymerase
Distilled water up to 50 µl

2.2.5 Cleavage of DNA

Restriction endonucleases (RE) are enzymes able to cut DNA molecule into pieces. These enzymes are found in bacterial cells where they function as a part of cell defence mechanism in which the restriction enzymes hydrolyses any exogenous DNA that appear in the cell. There are three types of restriction enzymes (I, II and III). The most commonly used in cloning are type II restriction endonucleases, these enzymes are able to recognize a specific target sequence and cut at a specific band relative to its target generating specific fragments (Dale, 2002, Reece, 2004).

Bovine serum albumin (BSA) is a large globular protein that has several biochemical applications. In restriction digests, BSA is used to stabilize some restriction enzymes during the digestion of DNA as well as to prevent the adhesion of enzyme to the reaction tube or any other vessels (NEB, 2008).

Reaction mixture
DNA 3-17 µl
10 x BSA 0.5 µl if needed (NEB)
Restriction enzyme 0.5 µl
10 x Buffer 2 µl
Distilled water to 20 µl

2.2.6 Separation of DNA on Agarose Gel Electrophoresis

A basic tool of separation and fractionation of nucleic acid is based on their size is gel electrophoresis. A charged molecule as DNA (negative charge) placed in an electric field will migrate toward the opposite electrode (the positive), smaller molecules migrate faster due to their light weight and smaller mass as well as because they face less frictional drag in the gel. The technique is based on the size of molecule not the charge that the molecule carries, since all the nucleic acids carry the same amount of charge per unit (Nicholl, 2002).

Procedure

Agarose gel (0.8%) was prepared by dissolving agarose in 1xTAE-buffer and allow the agarose to dissolve by bringing the solution to boiling point. The solution was poured into a gel chamber. The reaction mix 5-20 µl was mixed with 1/10 of its volume with loading dye in order to give the DNA both thickness and colour then amplified into the gel. A DNA marked was amplified into the gel as a standard.

2.2.7 Purification of DNA

The purification of DNA was carried out using the QIAquick procedures; the system combines the convenience of spin-column technology with the selective binding of silica-gel membrane. The DNA is adsorbes to the silica-membrane in the presence of high salt while contaminants pass through the column. Impurities were washed away and the pure DNA was

eluted with buffer or water. The procedures are carried out through tree steps bind, wash and elute (QIAGEN, 2008).

The QIAquick system is used for rapid DNA cleanup which includes a direct purification of PCR product (100 bp-10 kb), oligonucleotides and DNA up to 10 kb from enzymatic reactions (e.g. labelling, dephosphorylation, restriction and tailing) as well as extraction of DNA fragments from agarose gels (QIAGEN, 2008).

2.2.7.1 Purification of DNA from solution

Purification of DNA from solution was used to purify single- or double-stranded DNA fragments from PCR reactions. Fragments ranging from 100 bp to 10 kb were purified from primers, nucleotides, polymerases, and salts (QIAGEN, 2008).

Procedure

To 1 volume of PCR sample, 5 volume of buffer PB was added and mixed. The QIAquick spin column was placed in a 2 ml collection tube, the sample was applied and centrifuged for 1 min at 13000 rpm. The flow-throught was discarded and the column was placed back and the sample centrifuged at 13000 rpm for an additional 1 min.
The spin column was paced in a clean 1.5 microcentrifuge tube and the DNA was eluted adding 30 µl elution buffer (0.5 x T4-DNA ligation buffer to ligate afterward) to the center of the QIAquick membrane. The tubes then was centrifuged at 13000 rpm for 1 min (QIAGEN, 2008).

2.2.7.2 Purification of DNA from gel

Purification of DNA from gel to extract and purify DNA of 70 bp to 10 kb from standard or low-melt agarose gels in TAE buffer using *QIAquick gel extraction kit* protocol in a microcentrifuge (QIAGEN, 2008).

Procedure

The excised DNA fragment from the agarose gel was weighed and 3 volume of buffer QG was added to 1 volume of gel. The gel slice with QC buffer was incubated at 55°C for 10 min or until the gel slice has been completely dissolved. One volume of isopropanol was then added to 1 volume of gel (this step increases the yield of DNA fragment <500 bp and > 4 kb). The DNA sample was applied to the QIAquick column, the QIAquick spin column placed in a 2 ml collection tube and centrifuged for 1 min at 13000 rmp. The flow-through was discarded and the column was placed back to the collection and 750 µl of buffer PE was added. The sample was centrifuged for 1 min (13000 rpm), the flow-through was discarded and the QIAquick column was placed back and centrifuged at 13000 rpm for an additional 1 min. The QIAquick was placed into a clean 1.5 ml microcentrifuge tube and the DNA was eluted from the column into sterile eppendorf tube using 30 µl elution buffer (05.x T4-DNA ligation buffer) applied to the center of the QIAquick membrane, 1 min waiting and then was centrifuged at 13000 rpm for 1 min (QIAGEN, 2008).

2.2.8 Ligation

Ligation is joining two DNA fragments. The process of ligation or joining of DNA is carried out by an enzyme known as *DNA ligase*. DNA ligase is an enzyme which is able to repair single-strand breaks (nicks) in the sugar-phosphate backbond of the double-stranded DNA molecule. The action of ligase requires an exposed 3'OH group and a 5'-phosphate. There are two types of DNA ligases, the first uses NAD^+ as a cofactor and found in bacteria and the second uses molecule ATP as a cofactor and is fond in eukaryotes, viruses and bacteriophages.
DNA ligases such as T4 DNA ligase (encoded by the bacteriophage T4) are able to ligate blunt-ended DNA fragments (Dale, 2002). T4 DNA ligase enzyme works best at 37°C but is used at 4-15°C to avoid thermal denaturation of short base paired regions that hold the cohesive ends of DNA molecule together (Nicholl, 2002).

Procedure

The purified DNA from agarose gel or enzymatic reaction mixed in a total volume of 17 µl (13 µl of insert and 4 µl of vector). A 2 µl of 10 x ligase buffer added and 1 µl of ligase enzyme (e.g T4-DNA ligase). The mixture incubated overnight at 4°C the ideal temperature ranges from 9 to 14°C.

2.2.9 Gene transfer

2.2.9.1 Transformation

Bacteria exchange genetic information in both nature and laboratories, where one of these mechanisms is transformation. In transformation the bacteria takes up isolated DNA molecule form its surrounding medium; such prokaryotic cells are known to be competent. Mostly used prokaryote host in laboratories is *E. coli* K-12. The transformation can occur naturally in some cells, while other cells transformation should be carried under special conditions as $CaCl_2$ and heat pulse where only a small fraction will take up the plasmid DNA, typically 1 cell in about 10.000 takes a single plasmid and become transformed (Lodish, 2008, Klug *et al.*, 2006).

General laboratory transformation uses prepared competent cells, stored at -80°C. Cells are gently mixed with the DNA (plasmid DNA), and subjected to heat-shock. The cells then incubated in rich medium (e.g. SOC) for 1 h. Transformed cell are selected on plated with antibiotics.

2.2.9.1.1 Preparation of RbCl-competent cells

1% of an *E. coli* strain was inoculated from an over-night culture in Psi-medium (1 ml culture to 100 ml medium) at 37°C shacking incubator, until they reached an OD_{600} of 0.4. The measurements were carried out using spectrophotometer with Psi as blank. The cultures that have reached the correct OD were incubated on ice for 15 min. The cells were harvested by centrifugation at 4500 rpm for 5 min., and the supernatant were discarded and the cells were resuspended in 40 ml cold TFB1. The cells were incubated in ice for 5 min and were harvested again by centrifugation as in the previous step. The cells were resuspended in 3 ml

of cold TFB2. 100 µl of cells were transformed into sterile 1.5 ml tubes and were placed on ethanol and dry-ice for immediately freezing. The cells were stored at -80°C (NEB, 2009).

2.2.9.1.2 *Heat shock transformation of RbCl-competent cells*

Competent cells (100 µl) were placed on ice, and 10 µl of ligation mix were mixed gently. The cells were incubated on ice from 30 min-1 h. The cells were subjected to heat shock at 37°C for 1 min and were incubated back on ice for 2 min. Preheated 900 µl of SOC medium was added to the cells and incubated in shacking incubator at 37°C for 1-2 h. 200 µl of transformed cell were plated on LA with antibiotics for selection. The rest of cells were harvested by centrifugation for 1 min at 4500 rpm, 600 µl of cells were discarded and the rest (200 µl) were resuspended and were plated on LA with antibiotics as a concentrated sample.

As negative control competent cells with no DNA added was used to test for contaminations and selection plates. The positive control only 1 µl of plasmid DNA transformed and plated on LA with antibiotics. The positive control is used to test the efficiency of the competent cells and transformants (NEB, 2009).

2.2.9.2 Conjugation

Another way of gene transfer is conjugation. The conjugation is a natural process in some bacteria where the genetic information is transferred from one bacterium to another horizontally. The process was firstly studied by Lederberg and Tatum in 1946. The physical and genetic bases of the conjugation, where one cell serve as donor of parts of its chromosome and designated as F^+ cells and receiving cells that recombine the received DNA as a part of its own genome, they are designated as F^-cells. During the conjugation process the donor cell and the recipient cell are connected to each other through a pilus. Unlike transformation a cell-to-cell contact is required. There are some other requirements for the process as the origin of transfer the *oriT* and pilus proteins (Klug et al., 2006).

In this work the donor cells for conjugation were *S 17.1 λ pir* carrying the designated plasmids of interest. The *S 17.1 λ pir* are *E. coli* cells derived from *S17.1* strains. The *S 17.1 λ pir* are a lysogenic strain containing a λ pir as it codes for π protein which is required for the replication of R6K-ori (de Lorenzo *et al.*, 1993).

Procedure

Recipient cells *A. vinelandii* ATCC12158 were inoculated in BM for 3 days at 30°C shaking incubator. The donor cells were inoculated in LB with appropriated antibiotics at 37°C in shaking incubator over night. 1% of recipient cells culture was inoculated in new medium for 20 hours, as well as 1% of donor cells culture was incubated in a new medium for 2 hours. Both donor and recipient cells were mixed 3 to 4 ml and were centrifuged for 10 min at 5000 x g, the supernatant was discarded and the remaining cells in 200 µl were resuspended. The resuspended cells were spotted on a LA plates without antibiotics overnight. The grown cells on the LA plates were removed by scarping with a sterile spatula and were transferred into a 1.5 ml tube with 1 ml BM and a dilution series up to 10^{-4} were made. From each dilution mixture it was plated 200 µl on BA plated with appropriated antibiotics for 3 days. The colonies were inoculated in BM at 30°C in shaking incubator for 3 days and stored or used further in selection.

2.2.9.2.1 Selection of transconjugantes

The selection method used in this work is based on the *sacB* gene from *B. subtilis.* The *sacB* gene encodes levansucrase. The gram-negative bacteria that express and activate the *sacB* gene in the presence of sucrose are producing levan that appears to be lethal (Pelicic *et al.*, 1996).
The integration of mutant gene (cloned in vector) to the bacteria strain is through homologous recombination results in replacement of the wild-type and the removal of the vector plasmid due to its inability to grow inside the cell autonomously (Jacobson *et al.*, 1989).

In this work the mutant version of the gene of interest was cloned into a suicide vector containing the *pir*-dependent R6K replicon, the *sacB* gene and the Sp resistant gene as selective markers (pHE216 and its derivative pMK12). After conjugation of vector plasmid to the recipient cells as described in subsection 2.2.9.2, two homologous recombinations take place in the first and the second single-crossover event resulting in plasmid integration followed by excision (Blomfield *et al.*, 1991).
The first single-crossover event results in the integration of the vector plasmid at homologous sites (not necessary identical sites). After the second single-crossover event the plasmid becomes excised due to its inability to replicate within the recipient cell, it becomes lost.
To provide a direct selection for the plasmid loss and to ensure that the vector was not integrated into the chromosome of the recipient cell, the transconjugantes are selected on sucrose 5%. The resulting segregrants are Sucrose^R colonies then screened for the loss of Sp^R to ensure that the sucrose-resistant phenotype is due to loss of the integrated plasmid (Blomfield et al., 1991).

The result after the two homologous recombinations is either (i) the wild type is reconstructed and the mutant is lost with the plasmid or (ii) the wild type is lost with the plasmid and the mutant is constructed in the chromosome (Blomfield et al., 1991).

Procedure

1% of the inoculated cells after conjugation in RA1 medium (see discussion), were inoculated in new RA1 medium with appropriated antibiotics at 30°C in shaking incubator over night. The process was repeated in 4 additional days. On the fifth day the 1% culture was inoculated in new RA1 medium without antibiotics to allow excision of vector (the vector to recombine out of chromosome). The dilution series were made from the fifth day culture up to 10^{-4}. From each dilution mixture it was plated 100 μl on 5% sucrose-RA1 plates for 2 days in 37°C incubator. The colonies from the sucrose plates were transferred using a sterile tooth pin to RA1 plates with antibiotics and with out antibiotics. Those colonies that were able to grow on the plates with no antibiotics and failed to grow on plates with antibiotics are selected as possible mutant strains. The selected transconjugantes were inoculated in RA1 medium at 30°C in shaking incubator for 2-3 days and used in PCR for confirmation (Gimmestad *et al.*, 2009).

2.2.10 Sequencing of DNA

2.2.10.1 Sequencing of PCR products and plasmids

Mutations can occur randomly during the PCR reaction. The importance in analysing is particularly the PCR products if they are going to become a part of cloning procedures, as its important to verify their sequence. Once the PCR product has been cloned into a vector, the recombinant molecule can be used for DNA sequence analysis of the inserted PCR. The general known method for sequencing is the Sanger method published in 1977. The method is based on synthesis of new DNA from primer annealing to a single-stranded template DNA molecule. Various DNA polymerase can be used in the sequencing mixture, as well as the four dNTP (dATP, dCTP, dGTP and dTTP) added to the reactions to synthesise DNA by extension from its 3'end. The ddNTP should be present in each of the four reactions as it lacks the 3'hydroxyl group (3'OH). The ddNTP is required to generate chain determination as the absence of the 3'OH prevents the formation of new phosphodiester bond. Thus the generated fragments from the sequencing reaction varies in length and are separated on high – resolution denaturing polyacrylamid gel, they will migrate according to their length (smaller faster). This migration will create a ladder of fragments for each base (e.g A) presenting its position in the DNA sequence. Comparing the migration rates from the different reactions allows the sequence of DNA to be created. In order to read the sequence, each ddNTP is labelled with different fluorescent dye. The sequence then can be red with help of laser (Reece, 2004, McPherson & Møller, 2006).

The sequencing reactions used in this work was BigDye Terminator Ready reaction Mix v1.1 (Applied Biosystems). Plasmids, genomic DNA and PCR products were sequenced in this work. The sequencing reaction runs on the PCR-devise from Eppendorf (Eppendorf Mastercycler Personal) with the following reaction conditions.

The sample was incubated for 4 min at 96°C.
30 sec for denaturation at 96°C, 15 seconds for anneling at 50°C and 4 min for elongation at 60°C. The cycles were repeated for 24 times. The sample cooled at 4°C.
The sample can be stored on ice for purification or stored at -20°C.

Reaction mixture
150-300 ng plasmid DNA, 10-40 ng PCR product
1 µl primer (each primer, 3.2 pmol/ul)
4 µl dilution buffer (250 mM Tris-HCl (pH9.0), 10 mM MgCl$_2$)
Distilled water to 20 µl (including the volume of BigDye)
4 µl BigDye Terminator Ready reaction Mix v1.1 (Applied Biosystems) added last.

Purification of sequencing product

The sequencing reaction mixture was transferred into 1.5 ml tube and was precipitated by the addition of 2 µl NaAC (3 M, pH 5.2) and 50 µl of 96% ethanol. The sample was mixed well and was kept at room temperature for 15 min. The sample was centrifuged at 13 000 rpm for 20 min. The supernatant was removed and the sample was washed with 250 µl of 70% ethanol. The sample was mixed well and was centrifuged for 5 min at 13 000 rpm. All the supernatant was removed and the sample was allowed to dry with open lid on bench over night. The sample was delivered for separation of the sequencing product.

2.2.11 Protein purification and analysis

2.2.11.1 Preparing of cell extract for protein study

Bacterial cells were inoculated in 10 µl 3x LB with appropriated antibiotics, overnight at 37°C in shaking incubator. 500 µl of the overnight culture was transferred into new 500 ml flask with 50 ml 3x LB and antibiotics and incubated at 30°C shaking incubator for 3 hours. m-toluic acid (final concentration 0.5 mM) was added as an inducer for *Pm*-promoter and the culture incubated for 4 additional hours. The culture was transferred into 50 ml plastic tubes. The volume and OD_{600} was estimated. The cells were harvested by centrifugation for 10 min at 5000 x g and the supernatant was discarded .Cell's pellet was stored at -80°C for further use.

All the work with proteins was carried out on ice. The cell free extract was resuspended in 1/10 of its initial volume before harvesting in 50 mM final concentration of Tris-HCl (pH 7.7) and the cell were disrupted by sonication for 10 min with 8 sec. interval. The cells were centrifuges at 6500 x g for 45 min to remove the cell debris, followed by filtration in 0.2 µm. The cell free extract was used further in the enzyme assay and for protein purification (subsection 2.2.11.2 and 2.2.11.3).

2.2.11.2 Transketolase assay

Transketolase assay was based on the assay from **Sigma-Aldrich** with some modifications based on transketolase assay described in Iida et al 1993.

The mixture for enzyme assay up to 1 ml in each cuvette is shown in page 44. The activity of transketolase was measured spectrophotometrically by measuring the reduction in NADH every 10 min with interval of 10 sec. at 37°C, and 340 nm with *xylulose 5-phosphate* and *ribose 5-phosphate* as substrates (Iida et al., 1993, Sigma-Aldrich, 1994).

The solution mixtures were incubated on 37°C for 40 min prior to measuring the activity in order to stabilize the substrates. The enzyme was added right before measuring the protein activity. The Blank mixtures contain no enzyme (Sigma-Aldrich, 1994).

2.2.11.3 Protein expression and size determination using SDS-PAGE

SDS-PAGE (sodium dedocyl sulphate-polyacrylamide gel electrophoresis) is used to separate protein by size. The proteins samples are treated by boiling with SDS (denatured), in order to unfold the polypeptide chain and coat the entire strand of amino acids with a negative charge.

The polyacrylamid gel has small pores that the gel electrophoresis used to separate DNA fragments thus is suitable for protein separation as proteins are smaller than DNA molecules. When the electric current is applied the proteins will migrate away from the negative pole, where smaller proteins migrate faster and circumnavigate easier through the pores of the acrylamide gel. Because the charge-to-mass ratio is nearly the same among denatured proteins by SDS, and the final separation of proteins is dependent mostly on the differences in relative molecular mass of polypeptides (Clark & Pazdernik, 2009).

The loading buffer (chapter 2.4), contains; **bromophenol blue (BPB)** which is a dye used to visualize how the sample moves in the gel. BPB is visualized in natural pH and alkali. The **glycerol** is used as thicking agent as it helps the protein sample to be loaded in the wells. **The Dithiotheritol (DTT)** is a reducing agent used to disrupt disulfide bonds to allow full

denaturation of proteins before loading on the gel as well as to ensure that the proteins run uniformaly.

The separated proteins then are visualized using Coomassie brilliant blue (dark blue dye) that binds tightly to all proteins (Clark & Pazdernik, 2009).

Procedure

Clear PAGE ^{TM} SDS-polyacrylamide gel (10%), 10x10 prepared by C.B.S Scientific Company Inc. was used in this work. The polyacrylamid gel was placed in the chamber. The running buffer (see solutions and mediums) was poured into a gel chamber. The cell free extract was prepared as shown in chapter 2.1.11.1 and was mixed with 3/10 of its volume with 4X SDS-loading buffer (see solutions and mediums). The mixed samples were boiled at 100°C for 5 min. and were loaded into gel wells. Gels of 12 wells are capable of loading samples up to35 μl, and of 17 wells are capable for loading samples up to 17 μl. A protein ladder (10-230 kb) of 5 μl was amplified into the gel as a standard. The gel was run for 1-1.5 h, and was then stained using protein staining buffer for 1 h. The destaining of the gel was carried out for 10 min in destaining buffer. The destaining buffer was poured off and the gel was incubated in a new destaining buffer under continues shacking overnight. The washing procedures were repeated for some additional times until the dark blue dye washed away and the gel was visualized.

2.2.11.4 Protein purification using Column Chromatography

Chromatography is several techniques where in all the sample of the molecules, the analyte is dissolved in a mobile phase and run through the stationary phase. There are several types of FPLC-High Pressure Liquid Chromatography where the ion – exchange chromatography is one of them. The mobile phase is a low salt concentration solution thus low conductive, in order to optimize the binding of all charged molecules. The adsorption of the molecules to the solid support is derived by the ionic interaction between the oppositely charged molecule and the ligand in the support. As the salt concentration increased, the molecules of weakest ionic interaction will be eluted first from the column and those of higher or stronger ionic strength will be eluted later. The separates at different fractions are being collected at the column exit. As the mobile phase exits the column, the detector emits and draws peaks on the chromatogram. The emition of the detector results from the response to molecules in the eluting sample (Amersham, 2003, Clark & Pazdernik, 2009).

Procedure

The pumps of chromatography instrument were firstly washed with the two buffers A and B for column chromatography (see solutions and mediums) and then a washing program for columns was run to remove unbound materials.

The cell free extract was prepared from 250 ml culture as described in subchapter 2.2.11.1 and was mixed with ½ of it volume with cold Tris-HCl (50 mM final concentration) pH 7.7 and loaded into the 80 ml glass column of the chromatography device with flow rate of 4ml/min and 280 UV.

The loaded sample was collected as 10 ml fractions into 15 ml tubes. The fractions were frozen at -20°C for further use in enzyme assay or/and SDS-PAGE.

2.2.11.5 Estimation of protein concentration using Bio-Rad standard assay

The soluble protein concentration was estimated using Bio-Rad assay, by adding an acidic dye and measuring at 595 nm with spectrophotometer. The standard curve obtained provides a relative measuring of protein concentration.

The principle of the Bio-Rad Assay is based on dye-binding where the change of dye color occurs in response to protein concentration. The dye reagent Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when the binding to proteins take place (Bio-Rad, 1994).

Procedure

Six dilutions of the standard protein BSA (0, 1, 2, 4, 6 and 8 µg/ml) were prepared. 800 µl of BSA protein standard was pipetted into new sterile 1.5 ml Eppendorf tubes and the 200 µl reagent dye was added. The 200 µl reagent dye was also added to 800 µl of distilled water and to 800 µl of protein. The protein samples were prepared in triplicate. The samples were vortexed and incubated at room temperature for 20 min. The absorbance was measured at 595 nm at room temperature (Bio-Rad, 1994).

2.3   Bioinformatic tools

2.3.1 NCBI/BLAST

BLAST states for *Basic Local Alignment Search Tool*, the program was developed at the National Center of Biotechnology Information by Stephen Atschul and others in 1990. BLAST is the most used database search program (Ignacimuthu, 2005, Zvelebil & Baum, 2008).

BLAST is an algorithm for comparing biological sequences, such as amino acid sequences obtained from different proteins or DNA sequences. BLAST lies in comparing a query sequence (our sequence of interest) with a library or database of sequences, and identifies a library sequences that look like or have the similar appearance to the query sequence (Cristianini & Hahn, 2007).

There are different BLAST programs such as BLASTP which compares amino acid query sequence against a protein sequence found in the database. BLASTN compares nucleotide query sequence against a nucleotide sequence database. BLASTX compare six-frame translation products of nucleotide query sequence against a protein sequence database (Ignacimuthu, 2005).

2.3.2 Clone manager

Clone Manager is a program from Scientific and Educational Software. Clone Manager can basically be used to help with cloning also to be used to identify ORFs, primer design, identify potential start codon (ATG and GTG), and translate genes as well as for graphic map drawing (Sci-Ed, 2007).

2.3.3 ClustalW

ClustalW is a bioinformatics tool that uses a tree method of multiple alignment. The program available and has access from a number of web sites. ClustalW works in a way that person should collect wished sequences to be studied in form of either DNA or protein sequences. The uploaded sequences will be paste into a dialog box and the program runs. It is possible to view the aligned sequences in colours as well as the phylogeny tree can be displayed (Zvelebil & Baum, 2008).

2.3.4 Prosite, Interpro and pfam

Many proteins contain compact units within the folding pattern of a single chain, these are called domains. Proteins may have one or several domains which gives information about the function. There are different databases that can be applied in order to determine domains/motifs and a protein family (such as PROSITE and Pfam). InterPro is a database tool allows efficient searching, as well as it is an integral annotation resourse for **protein families**, **domains** and **functional sites** that combines several databases effort such as PROSITE (Thornton *et al.*, 2003, Ignacimuthu, 2005).

2.4   Solutions and mediums

- **<u>Burks medium (BM)</u>**

5x Burks Buffer

1 g/l $MgSO_4$ x $7H_2O$
3.2 g/l K_2HPO_4
0.8 g/l KH_2PO_4
1 g/l NaCl
Distilled water

1x Burks Buffer

200 ml of 5x Burks Buffer
700 ml of Distilled water

The 1x Burks Buffer was autoclaved for 20 min. at 120°C. The buffer stored at room temperature

Salt solutions

21 mg /ml Na_2MoO_4 x H_2O
0.5 g/ml CaCl x 2 H_2O
150 mg/ml $FeSO_4$ x 7 H_2O

Salts were dissolved in distilled water, sterilized by filtration and stored at the -20°C freeze.

Carbon source

2% of sugars (Glucose and sucrose were used in this work)
20 g/l of glucose was dissolved in 1 L water and autoclaved. The solution was stored at room temperature.

Burks Medium

The ingredients are autoclaved or sterile filtered and mixed as listed;

900 ml of 1x Burks Buffer
100 ml carbon source
100 µl of each salt solution

Burks agar

15 g/l of agar added to 1x Burks medium

The medium autoclaved for 20 min. at 120°C, cooled to 50°C in water bath and then the carbon source, salt solutions and the antibiotics are added. The medium is poured on Petri dishes and stored in refrigerator at 4°C for around 6 months.

- ## **Luria-Bertani (LB) medium**

10 g/l trypton
5 g/l yeast extract
5 g/l NaCl
Distilled water

The medium autoclaved for 20 min. at 120°C.

3x LB

30 g/l trypton
15 g/l yeast extract
5 g/l NaCl
Distilled water

The medium was autoclaved for 20 min. at 120°C. The medium is stored at room temperature.

LA medium

LB-medium with 15 g/l agar

The medium autoclaved for 20 min at 120°C, cooled up to 50°C in water bath then the antibiotics added. The medium poured in Petri dishes and stored in refrigerator.

- ## **Psi medium**

20 g/l trypton
5 g/l yeast extract
5 g/l MgSO4 x7H2O
Distilled water

The solution is pH adjusted to 7.6 using KOH, autoclaved for 20 min. at 120°C. The solution is stored at room temperature.

TFB1

2.94 g/l KAc
12.0 g/l RbCl
1.5 g/l $CaCl_2$ x$2H_2O$
 10 g/l $MnCl_2$x $4H_2O$
150 ml/l glyserol
Distilled water

The solution pH adjusted to 5.8 using acetic acid. The solution sterilized by filtration and stored in the refrigerator.

TFB2

2 g/l MOPS
1.2 g/l RbCl
11 g/l $CaCl_2$ x$2H_2O$
150 ml glycerol
Distilled water

The solution pH adjusted to 6.5 using NaOH. The solution sterilized by filtration and stored in the refrigerator.

- ## RA1 medium

Salt solutions

147.02 g/l $CaCl_2$
50 g/l K_2HpO_4 pH adjusted to 7.0 (phosphate solution).

TMS1

5 g/l $FeSO_4$ x $7H_2O$
0.39 g/l $CuSO_4$ x $5H_2O$
0.44 g/l $ZnSO_4$ x $7H_2O$
0.15 g/l $MnSO_4$ x $2H_2O$
0.010 g/l Na_2MoO_4 x $2H_2O$
50 ml/l concentrated HCl
Distilled water

The solution is sterile filtered and stored at room temperature, away from day light.

RA1 medium basic solution

2 g/l $MgSO_4$ x $7H_2O$
1.5 g/l NH_4NO_3
2 g/l Pepton
10.5 g/l MOPS

Medium dissolved in 900 ml distilled water and the pH adjusted to 7.0 using 5-10 M NaOH. The medium is autoclaved for 20 min at 120°C and stored at room temperature.

RA1 medium

900 ml RA1 medium basic solution (autoclaved)
2 ml CaCl2-salt solution

100 ml K_2HpO_4- salt solution
3 ml TMS1
100ml/l of sugar solution (2% glucose) prepared and autoclaved separately.

The medium can be stored at room temperature.

RA1-agar medium

15 g/l agar is added to RA1-medium basic solution and autoclaved. The medium cooled to 50°C in water bath and then the $CaCl_2$, K_2HpO_4, TMS1, carbon source and the antibiotics are added. The medium poured into Petri dishes and stored in the refrigerator for 6 months.

- ## SOC medium

20 g/l trypton
5 g/l yeast extract
0.5 g/l NaCl
3.6 g/l glucose
5.08 g/l $MnCl_2$ x2H_2O
2.5 mM KCl
Distilled water

The medium sterile filtered and poured into 1.5 ml tubes. Medium sterilized by filtration and stored at -20°C.

- ## Stock solutions for antibiotics and mtolic acid

Ampecilline (Ap)

200 mg/ml ampicilline dissolved in distilled water. The solution sterilized by filtration and stored at -20°C. Used in the medium as 200 µl /ml.

Tetracycline (Tc)

15 mg/ml tetracycline dissolved in 50% ethanol. The solution sterilized by filtration and stored at -20°C. Used in the medium as 15 µl /ml.

Spectinomycine (Sp)

20 mg/ml spectinomycine dissolved in distilled water. The solution sterilized by filtration and stored at -20°C. Used in the medium as 20 µl /ml.

1M m-toluic acid

136 g/mol dissolved in 96% ethanol, the solution was sterilized by filtration and stored in 1.5 ml tubes at -20°C.

- ## **Stock solutions for transketolase assay**

1M Tris-HCl (7.7pH) buffer

12.1 g Trizma-base was dissolved in 80 ml water and pH adjusted to 7.7 with concentrated HCl. Distillate water was added to total volume of 100 ml. The solution was autoclaved and stored at room temperature.

100 mM xyloluse 5-phosphate

5 mg dissolved in 198 μl distilled water and stored in -20°C.

50 mM ribose 5-phosphate

13.70 mg of ribose 5-phosphate dissolved in 1 ml distilled water and stored at -20°C.

0.10% (w/v) Thiamine pyrophosphate (TPP)

0.001 g of Thiamine pyrophosphate was dissolved in 1 ml. The solution prepared fresh, not storable.

300 mM MgCl$_2$

6.009 g of MgCl$_2$ was dissolved in 100 ml of distilled water and autoclaved. The solution stored at room temperature.

NADH

0.0035 g of NADH was dissolved in 1 ml distilled water. The solution prepared fresh before use.

a-Glycerophophate Dehydrogenase/Triosephosphate Isomerase Enzyme Solution(a-GDH/TPI)

Delivered from Sigma-Aldrich as 2000 units/mg used as 2000 units/ml.

Table 2.4 shows the final concentrations prepared to 1ml total volume for transketolase assay.

Final concentration	Volume used to total 1ml per sample
50 mM of Tris-HCl (7.7pH)	50 μl of 1M Tris-HCl (7.7pH)
0.5 mM of Ribose 5 phospahte	10 μl of 50mM Ribose 5 phosphate
0.5 mM of Xylose 5 phosphate	5 μl of 100mM Xyloluse 5 phosphate
0.10% of TPP	12.5 μl
300 mM MgCl$_2$	37.5 μl
4.3 mM NADH	25 μl
2000 unit/ml a-GDH/TPI	2.5 μl
Cell extract	100 μl
Distilled water	575.5 μl

- ## **Buffers for SDS-PAGE**

  ### **SDS-loading buffer**

4 x SDS-loading buffer is modified from 1x SDS loading buffer in molecular cloning 2. The volume prepared is 5 ml of 4x SDS- loading buffer.

Table 2.5 shows the final concentrations prepared to 5 ml total volume of 4x SDS-loading buffer for SDS-PAGE.

Final concentration	Volume mixed to 5 ml total
200 mM Tris-HCl (6.8pH)	1 ml of (1M) Tris-HCl (6.8 pH)
400 mM DTT	2 ml of (1M) DTT
8% SDS	0.4 g SDS powder
0.4% Bromophenol blue	0.2 g
40% glycerol	2 ml of (100% glycerol)

The volumes mixed and water added up to 5 ml, the buffer used immediately or stored at -20°C.

SDS-running buffer (x1)

3.02 g Trizma-base
18.8 g Glycine
5 ml 20% SDS
900 distilled water

Mixed and water added to total volume of 1 liter.

SDS-distaining solution (methanol-acetic acid)

500 ml methanol

400 ml distilled water
100 ml galacial acetic acid

Mixed and stored at room temperature.

SDS-staining solution

In 100 ml of methanol-acetic acid solution, dissolve 0.25 g of Commassie Brilliant Blue
Mix well and store at room temperature. The staining solution can be used for several times.

- ## Buffers for column chromatography

Buffer A

50 mM final concentration of Tris-HCl (7.7pH) diluted from 1 M Tris-HCl (7.7 pH) described
in page 43, to total volume of 1 L. The buffer was sterilized by filtration and stored at 4°C.

Buffer B

50 mM final concentration of Tris-HCl (7.7 pH) with 1 M final concentration of NaCl was
prepared by diluting 1 M Tris-HCl (7.7 pH) and 5 M NaCl to total volume of 1 L. The buffer
was sterilized by filtration and stored at 4°C.

- ## Buffers and solutions for isolation of genomic DNA from *A. vinelandii*

TE (pH 8.0)

10 ml/l 1 M Tris-HCl (pH 8.0) autoclaved seperatly. Preparation of 1 M Tris-HCl is described
previously.
2 ml/l 0.5 M EDTA (pH 8.0) autoclaved seperatly. 0.5 M EDTA prepared by dissolving
186.12 g EDTA per 1 L dH$_2$O.
Distillate water up to total volume of 1 L.

0.9% NaCl, 10mM EDTA

1.35 g of NaCl and 0.55 g of EDTA dissolved in 150 ml of dH$_2$O.

10% SDS

15 g of SDS was dissolved in 150 ml of distilled water. The solution stored at room
temperature.

5M NaCl

43.83 g of NaCl dissolved in 150 ml of distilled water. The solution was autoclaved and
stored at room temperature.

0.72 M MgCl₂

21.95 g of MgCl2 was dissolved in 150 ml of distilled water. The solution was autoclaved and stored at room temperature.

3 Results
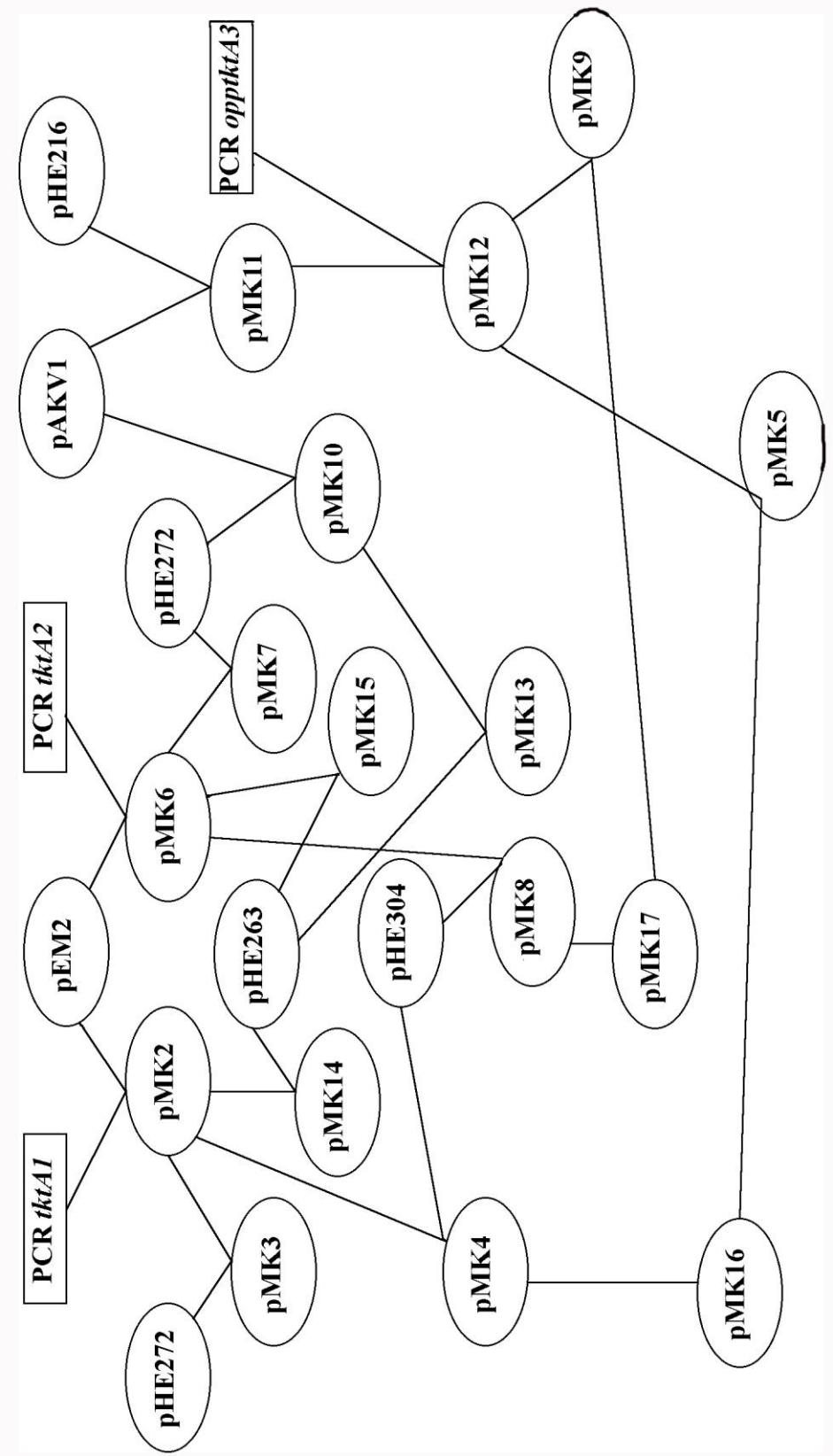
Figure 3.1 Scheme for the constructed vector plasmids used in this work.

The results are divided into three main parts (i) construction of plasmids for gene inactivation (3.2) and transposon vectors for complementation (3.1), (ii) inactivation of the three genes through conjugation to *A. vinelandii* strain (3.3) (iii) expression of three transketolase genes in *E. coli* (3.4).

Three transketolase genes *tktA1*, *tktA2* and *tktA3* were inactivated in three different vector plasmids pMK5, pMK9 and pMK12 respectively through conjugation to *A. vinelandii* ATCC12518 to study their influence on the alginate production.

The three mutant genes (*tktA1⁻*, *tktA2⁻* and *tktA3⁻*) in *A. vinelandii* ATCC12518 were complemented to pMK3, pMK7 and pMK10 transposon complementation vectors containing the three wild-type transketolase genes *tktA1*, *tktA2* and *tktA3* respectively (Figure 3.1).

The three transketolase genes from vectors pAKV1 (*tktA3*), pMK2 (*tktA1*) and pMK6 (*tktA2*), were sub-cloned into over-expression vector (pHE263) under control of *Pm*–promoter and grown in *E. coli K-12* production strains in order to identify, characterize and carry out an expression analysis of the three gene products.

3.1   Construction of transposons for complementation

Three wild-type genes *tktA1*, *tktA2* amplified from the genomic DNA of *A. vinelandii* and *tktA3* (obtained by Vie A. K. 2008) each was sub-cloned in to a suicide Tn5 minitransposon vectors (pHE272) under control of the mutant *PmG5*-promoter (described in chapter 1.3) in order to complement the three mutant transketolase genes in *A. vinelandii.*

The modified Tn5 minitransposon, is unable to migrate further after it had reaches the chromosome due to the modification made to its transpoases, the protein required for the transposon to transpose further. The modified Tn5 minitransposon used to introduce the wild-type version of the gene into the chromosome of *A. vinelandii* as transposon can transpose and settle where else in the genome. Three transposon vectors for complementation were designated pMK3 (*tktA1*), pMK7 (*tktA2*) and pMK10 (*tktA3*). Figure 3.1 described the cloning strategy.

3.1.1 Cloning of *tktA1* and *tktA2*

3.1.1.1 Amplification of *tktA1* and *tktA2* with PCR

Genomic DNA was isolated from *A. vinelandii* and the wanted DNA-fragments were amplified using PCR primers fwd/-rev as shown in table 2.3 and the Pwo-super yield PCR kit. The *tktA1* and *tktA2* PCR products were analysed using gel electrophoresis and the expected fragments were obtained (Figure 3.2).
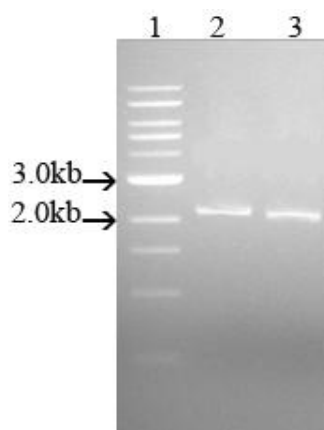
Figure.3.2 The PCR products separated by ethidium bromide stained agarose gel electrophoresis, the expected fragments are shown in parenthesis. Lane 1, 1 kb DNA ladder, lane 2, tktA1 (2.2 kb) and lane 3, tktA2 (2.2 kb).

3.1.1.2 Cloning of *tktA1* in pEM2

The PCR product of *tktA1* was digested using *NdeI* and *NotI* and purified using QIAquick kit. The vector plasmid pEM2 was digested with *NotI* and *NdeI* to give fragments of 3.9 kb and 1.2 kb (Figure 3.3a). The 3.9 kb vector fragment was ligated to *tktA1*. The ligation mixture was transformed into competent cells *E. coli S17.1 λ pir*. The transformed cells were selected on LA plate with Amp overnight at 37°C. Two colonies were inoculated in LB with Amp at 37°C shaking incubator for plasmid isolation. The isolated plasmids were digested with *NotI* and *NdeI* and expected fragments 3.9 kb and 2.2 kb were observed on the gel (Figure 3.3b, only one colony shown). The plasmid chosen for further work was designated pMK2.
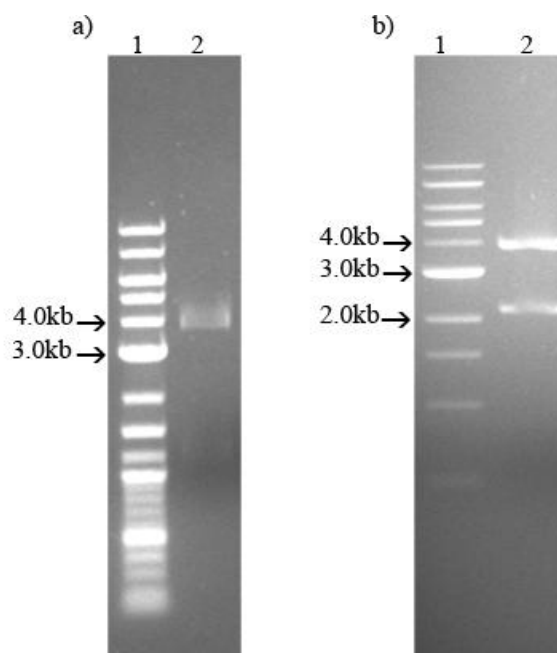
Figure 3.3 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, the expected fragments are given in parenthesis a) Lane 1, 2-log DNA ladder, lane 2, pEM2 digested with NotI and NdeI (3.9 kb and 1.2 kb)

3.1.1.3   Sequencing of *tktA1* in pMK2

During PCR amplification reaction mutations may occur. A change in the DNA sequence as a
result of mutation may affect the result by coding for another amino acid or introducing a stop
codon. To verify that the insert was as expected, the *tktA1* gene in pMK2 plasmid was
sequenced using the PCR-primers described in table 2.4. The alignment of *tktA1* in pMK2 and
the *tktA1* from *A. vinelandii* genome sequence showed that the nucleotide sequences of insert
obtained by PCR is 100% identical to the nucleotide sequence of *A. vinelandii* genome
(appendix H).

3.1.1.4   Construction of transposon vector pMK3 for complementation of *tktA1⁻*

Plasmid pMK2 was digested with *NotI* and *NdeI* and gave fragments 3.9 kb and 2.2 kb
(Figure 3.4a). The vector plasmid pHE272 was digested with *NotI* and *NdeI* to give fragments
of 8.0 kb and 1.7 kb (Figure 3.4b). The 2.2 kb fragment of pMK2 was ligated to the 8.0 kb
fragment of pHE272 and transformed to *E. coli S17.1 λ pir*. The transformed cells were grown
in LA with Amp overnight at 37°C. Two colonies obtained from two transformants plasmids
were isolated and digested with *NotI* and *NdeI*. The observed fragments were 8.0 kb and 2.2
kb as expected. Only one colony chosen for further studies and is shown in Figure 3.4c. The
generated plasmid was designated pMK3.
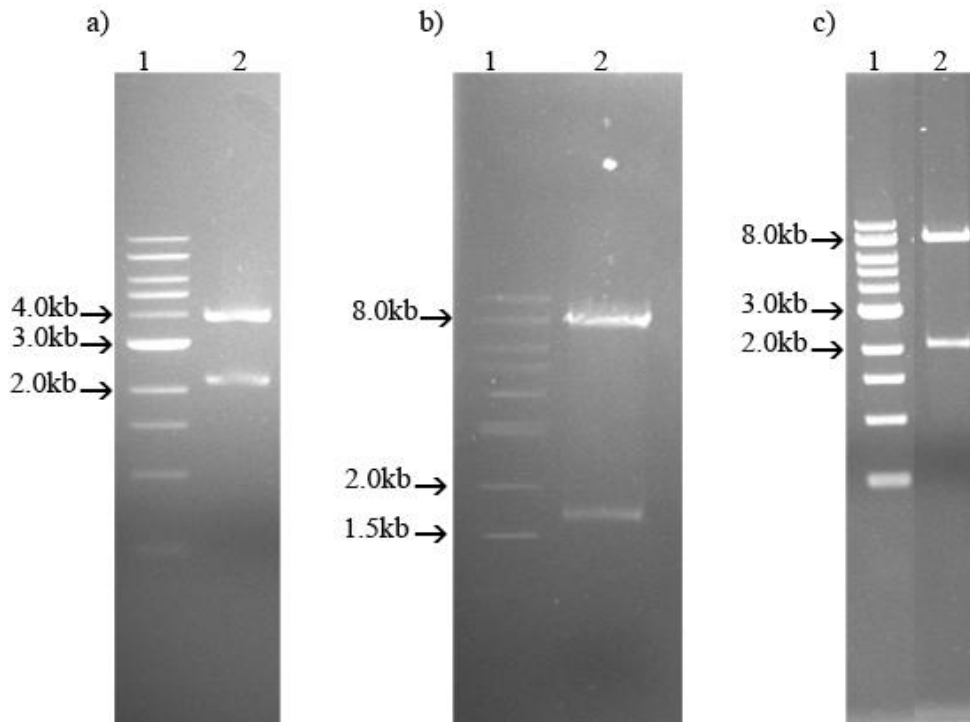
***Figure 3.4 The separated DNA fragments of digested plasmids by ethidium bromide
stained agarose gel electrophoresis,*** *the expected fragments are given in parenthesis **a)**
pMK2 digested with NotI and NdeI (3.9 kb and 2.2 kb).* ***b)*** *pHE272 digested with NotI and*

50

NdeI (8.0 kb and 1.7 kb). **c)** *pMK3 digested with NotI and NdeI (8.0 kb and 2.2 kb). The first lane in Figure 3.4a and b is 1 kb DNA ladder. The first lane in Figure 3.4c is 2-log DNA ladder.*

3.1.1.5   Cloning of *tktA2* in pEM2

The PCR product of *tktA1* was digested using *NdeI* and *NotI* and purified using QIAquick kit. (Figure 3.2). The pEM2 vector plasmid was digested with *NotI* and *NdeI* and two fragments obtained 3.9 kb and 1.2 kb (Figure 3.3a). The purified *tktA2* fragment was ligated to 3.9 kb fragment of the vector. The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells, and transformed cells were grown on LA plate with Amp overnight at 37°C. Two colonies were inoculated for plasmid isolation. The isolated plasmids were digested with *NotI* and *NdeI* and the expected fragments 3.9 kb and 2.1 kb were observed on the gel (Figure 3.5 only one colony shown). The plasmid was chosen for further work was designated pMK6.

Figure 3.5 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, the expected fragments are given in parenthesis, **Lane 1,** DNA 1 kb ladder, **Lane 2,** pMK6 digested with NotI and NdeI (3.9 kb and 2.2 kb).

3.1.1.6   Sequencing of *tktA2* in pMK6

To verify that the insert was as expected the *tktA2* gene in pMK6 was sequenced using PCR-primers described in table 2.4. The alignment of the nucleotide sequnces of the PCR-tktA2 and *tktA2* of *A. vinelandii* found to be 100% identical (appendix H).

3.1.1.7 Construction of transposon vector pMK7 for complementation of *tktA2*⁻

Plasmid pHE272 was digested with *NotI* and *NdeI*; two fragments were obtained 8.0 kb and 1.7 kb (Figure 3.4b). The pMK6 plasmid containing *tktA2* gene was digested with *NotI* and *NdeI* to give 3.9 kb and 2.1 kb (Figure 3.5). The insert of 2.1 kb obtained from pMK6 was ligated to the 8.0 kb vector fragment from pHE272, and transformed to *E. coli S17.1 λ pir* competent cells. The transformants were selected on agar plate with Amp. Two colonies were inoculated in LB with Amp overnight at 37°C for plasmid isolation. The isolated plasmids were controlled by digestion with *NotI* and *NdeI*, and gave two fragments as expected 8.0 kb and 2.1 kb are shown in Figure 3.6 (only one colony is shown). The transposon vector obtained was designated pMK7.
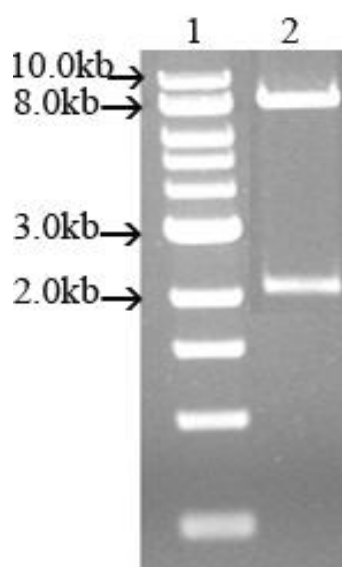
Figure 3.6 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are given in parenthesis,* ***lane 1,*** *DNA 1 kb ladder,* ***lane 2,*** *pMK7 digested with NotI and NdeI (8.0 kb and 2.2 kb).*

3.1.1.8 Construction of transposon vector pMK10 for complementation of *tktA3*⁻

The pHE272 vector plasmid was digested with *NotI* and *NdeI* to give 8.0 kb and 1.7 kb (Figure 3.4b), The 8.0 kb vector fragment was ligated to 2.2 kb of insert, obtained from pAKV1 (contains *tktA3* gene cloned by Vie, A. K. 2008) after digestion with *NotI* and *NdeI* (3.2 kb and 2.2 kb) (Figure 3.7a). The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. The four obtained colonies were inoculated in LB with Amp overnight at 37°C for plasmid isolation. The isolated plasmids were tested by digestion with *NotI* and *NdeI* and the expected fragments 8.0 kb and 2.2 kb were obtained Figure 3.7 b (only one colony is shown). The chosen transposon vector was designated pMK10.
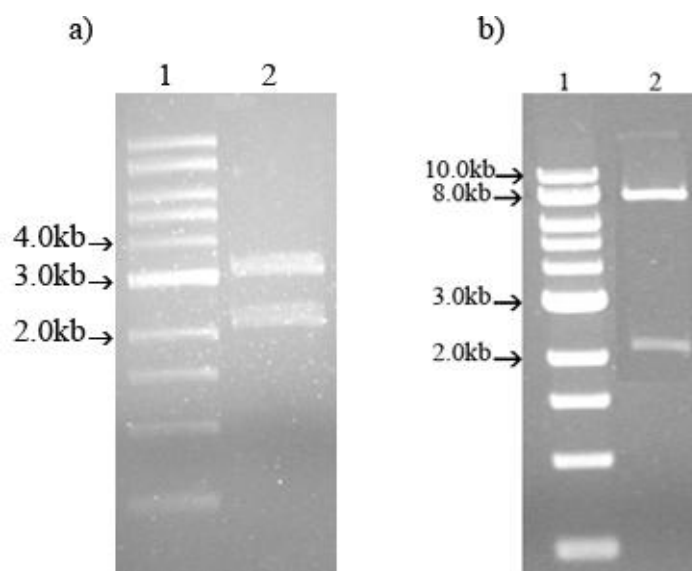
Figure 3.7 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are given in parenthesis **a)** pAKV1 digested with NotI and NdeI (3.2 kb and 2.2 kb) **b)** pMK10 digested with NotI and NdeI (8.0 kb and 2.2 kb). The first lane in both Figures 3.7a and b is 1 kb DNA ladder.*

3.2   Construction of plasmids for gene inactivation

To determine the *in vivo* function and influence of the three transketolase gene products on the synthesis of alginate it was developed a system where the three *tktA1, tktA2* and *tktA3* genes were disrupted by two single cross-over homologous recombinations (see subsection 2.2.9.2.1). Each mutant gene was designed by generating a deletion in frame and the flancking sequence was transferred into pHE216 vector plasmid and its derivative pMK12 containing *spectinomycin*[R] and *sacB* gene as selection markers, to generate pMK5, pMK9 and pMK12. The internal deletion in *tktA1*, *tktA2* and *tktA3* corresponds to deletion of amino acid residues 189 to 672, 189 to 672 and 1 to 440 in proteins respectively.
The designated plasmid vectors need π protein encoded by *pir* gene in the *E. coli S17.1 λ pir* for their replication. Thus the three mutant genes were designated in pMK5, pMK9 and pMK12 and transformed into *E. coli S17.1 λ pir* competent cells as a first host (intermediate strain).

3.2.1 Sequencing of *opptktA1/A2* in pHE303/pHE304

pHE303 and pHE304 were obtained by Helga Ertesvåg in 2009, by the ligation of *opptktA1* and *opptktA2* PCR products to pUC128. *opptktA1* in pHE303 and *opptktA2* in pHE304 were sequenced using the PCR-primers described in table 2.4.
The pairwise alignment of both PCR-opptktA1 to the *opptktA1* sequence designed in Clone Manager and the PCR-opptktA2 to the *opptktA2* sequence designed in Clone Manager program showed to be 100% identical (appendix H).

3.2.2 Sub-cloning of *opptktA2*

Only pHE304 containing the o*pptktA2* gene, was used further, as *opptktA1* from plasmid pHE303 has the *SacI* (GAGCTC) site which rendered the cloning procedures more difficult. The two genes of *opptktA1* and *opptktA2* differ in only one base pair located at 648 bp. The 648 base of *opptktA1* is guanine (G) while in *opptktA2* it is adenine (A). The difference in these base pairs creates the *SacI* site in *opptktA1* which is absent in *opptktA2* at that location. Thus plasmid pHE304 was digested with *SacI* and *PciI* to give three fragments 4.0 kb, 0.3 kb and 0.35 kb (Figure 3.8a). The plasmids pMK2 and pMK6 were digested with *SacI* and *PciI* and gave four fragments 2.6 kb, 1.3 kb, 1.1 kb and 0.8 kb (Figure 3.8a). The vector fragment of 4.0 kb from pHE304 was ligated to the 1.3 kb insert from pMK2 and pMK6 respectively. Each ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. The eight obtained colonies were inoculated in LB with Amp overnight at 37°C for plasmid isolation. The isolated plasmids from both transformations were tested by digestion with *SacI* and *PciI* and the expected fragments (4.0 kb and 1.3 kb) were obtained as shown in Figure 3.8b only one colony shown for each plasmid. The chosen plasmids were designated pMK4 and pMK8.
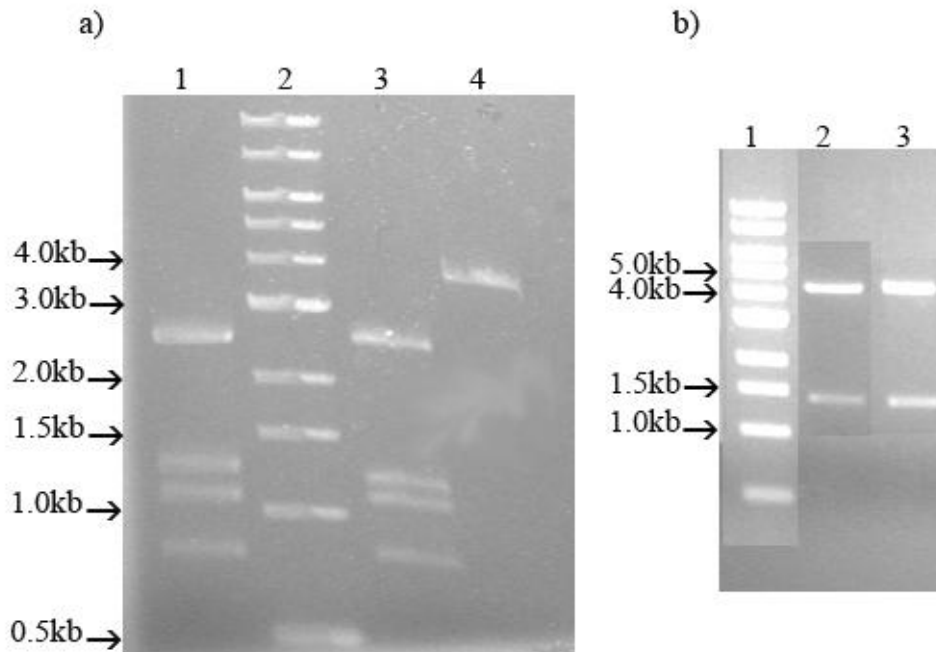
Figure 3.8 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are shown in parenthesis,* ***a) Lane 1,*** *pMK2 digested with SacI and PciI (2.6 kb, 1.3 kb, 1.1 kb and 0.8 kb),* ***lane 2,*** *DNA 1 kb ladder,* ***lane 3*** *pMK6 digested with SacI and PciI (2.6 kb, 1.3 kb, 1.1 kb and 0.8 kb) and* ***lane 4,*** *pHE304 digested with SacI and PciI (4.0 kb, 0.3 kb and 0.35 kb) due to small band sizes the 0.3 kb and the 0.35 kb is not shown.* ***b) Lane 1,*** *DNA 1 kb ladder,* ***lane 2,*** *pMK4 digested with SacI and PciI (4.0 kb and 1.3 kb),* ***lane 3,*** *pMK8 digested with SacI and PciI (4.0 kb and 1.3 kb).*

The pMK8 and pMK4 vector plasmids were digested with *SacI* and the obtained fragments were 5.2 kb and 5.1 kb for pMK8 and pMK4 respectively (Figure 3.9a). The digested fragments were religated and transformed into into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. The six obtained colonies were inoculated for plasmid isolation.

The isolated plasmids were tested by digestion with *AvrII* and *NotI-HF* and was expected to give two fragments of 3.0 kb and 2.2 kb for both the pMK4-deletion and pMK8-deletion as shown in (Figure 3.9b only one colony shown for each plasmid), the correct plasmids were obtained and designated pMK16 (pMK4-deletion) and pMK17 (pMK8-deletion).
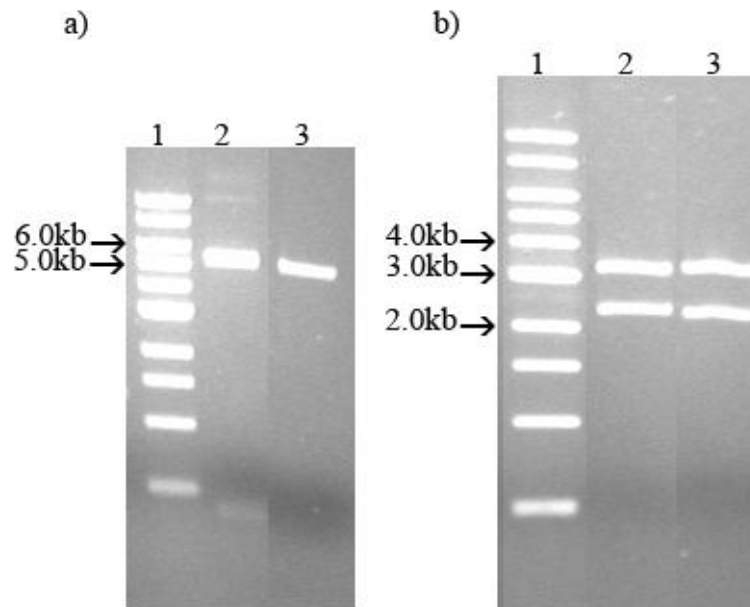
Figure 3.9 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are shown in parenthesis,* ***a) Lane 1,*** *DNA 1 kb ladder* ***lane 2,*** *pMK8 digested with SacI (5.2 kb),* ***lane 3,*** *pMK4 digested with SacI (5.1 kb).* ***b) Lane 1,*** *DNA 1 kb ladder,* ***lane 2,*** *pMK16 digested with AvrII and NotI-HF (3.0 kb and 2.2 kb),* ***lane 3,*** *pMK17 digested with AvrII and NotI-HF (3.0 kb and 2.2 kb).*

3.2.3 Construction of the *tktA1* gene inactivation vector pMK5

The pMK16 vector plasmid was digested with *SalI*, *NotI-HF* and *DraIII* to give four fragments of 2.5 kb, 2.0 kb, 0.5 kb and 0.1 kb (Figure 3.10b). The vector plasmid pMK12 was digested with *NotI* and *XhoI* and the obtained fragments were 11.4 kb and 2.0 kb (Figure 3.10a). The vector fragment of 11.4 kb was ligated to the 2.0 kb fragment of pMK16. The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. The two obtained colonies were inoculated in LB with Amp overnight at 37°C for plasmid isolation. The isolated plasmids were tested by digestion with *EcorV* and the expected fragments (8.5 kb and 4.9 kb) were obtained and shown in Figure 3.10c. The chosen plasmid for inactivation of *tktA1* gene was designated pMK5.
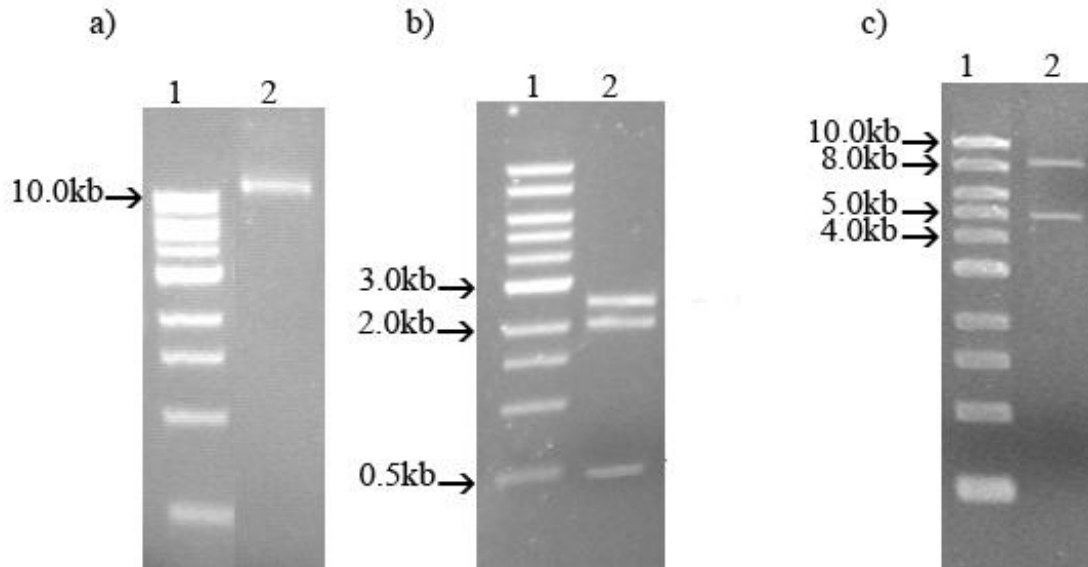
Figure 3.10 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, the expected fragments are shown in parenthesis a)Lane 1, DNA 1kb ladder, lane 2, pMK12 digested with NotI and XhoI (11.4 kb and 2.0 kb) due to low picture resolution the 2.0 kb band is not visible b) Lane 1, DNA 1kb ladder, lane 2, pMK16 digested with SalI, NotI-HF and DraIII (2.5 kb, 2.0 kb, 0.5 kb and 0.1 kb) the 0.1kb band is not shown due to its small size. c) Lane 1, DNA 1 kb ladder, lane 2, pMK5 digested with EcorV (8.5 kb and 4.9 kb).

3.2.4 Construction of the *tktA2* gene inactivation vector pMK9

The pMK17 vector plasmid was digested with *SalI*, *NotI-HF* and *DraIII* to give four fragments of 2.5 kb, 2.0 kb, 0.5 kb and 0.1 kb (Figure 3.11a). The vector plasmid pMK12 was digested with *NotI* and *XhoI* and the obtained fragments were 11.4 kb and 2.0 kb (Figure 3.10a). The vector fragment of 11.4 kb was ligated to the of 2.0 kb fragment of pMK16. The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. Only one colony was obtained and inoculated in LB with Amp overnight at 37°C for plasmid isolation. The isolated plasmids were tested by digestion with *EcorV* and the expected fragments (8.5 kb and 4.9 kb) were obtained and shown in Figure 3.11b. The plasmid for inactivation of *tktA2* gene was designated pMK9.
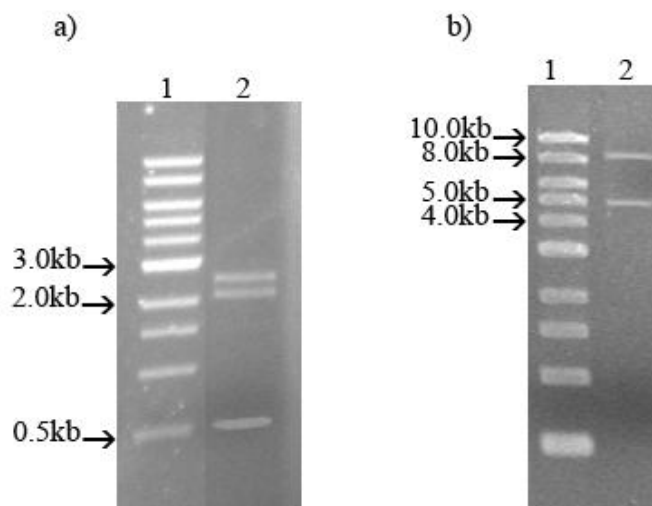
Figure 3.11 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, the expected fragments are shown in parenthesis a) Lane 1, 1kb DNA ladder, lane 2, pMK17 digested with SalI, NotI-HF and DraIII (2.5 kb, 2.0 kb, 0.5 kb and 0.1 kb) the 0.1kb band is not shown due to its small size.b) Lane 1, 1kb DNA ladder, lane 2, pMK9 digested with EcorV (8.5kb and 4.9kb).

3.2.5 Amplification and cloning of *opptktA3*

Genomic DNA was isolated from *A. vinelandii* and the DNA-fragments containing the sequence upstream of *tktA3* were amplified using PCR primers fwd/-rev as shown in table 2.3 were applied using Pwo-super yield PCR kit. The *opptktA3* PCR products were analysed using gel electrophoresis and the expected fragments were obtained (Figure 3.12).

The PCR products was analysed using gel electrophoresis and the expected fragment was obtained (1.7 kb).
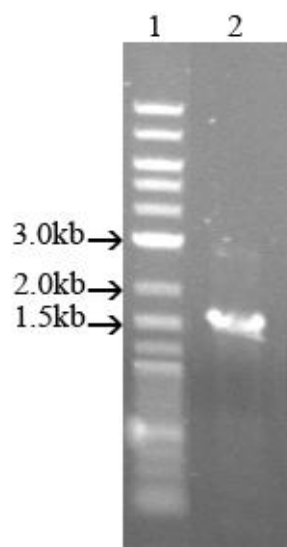
Figure 3.12 PCR products separated by ethidium stained agarose gel electrophoresis, the expected fragments sizes are given in parenthesis, lane 1,2-log DNA ladder, lane 2 opptktA3 (1.7 kb).

57

3.2.6 Cloning of *opptktA3* in pMK11

The PCR product of *opptktA3* was digested with *BglII* and *SpeI* and purified using QIAquick kit. The vector plasmid pAKV1 was digested with *BglII* and *SpeI* and two fragments obtained 4.1 kb and 1.3 kb (Figure 3.13). The purified *opptktA3* fragment was ligated to 4.1 kb fragment of the vector. The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells, and transformed cells were selected on LA plate with Amp overnight at 37°C. From two transformants plasmids isolated and digested with *SacI* and *PspOMI*. The expected fragments (3.1 kb and 2.6 kb) observed on the gel (Figure 3.14a only one colony is shown). The plasmid chosen for further work was designated pMK11.
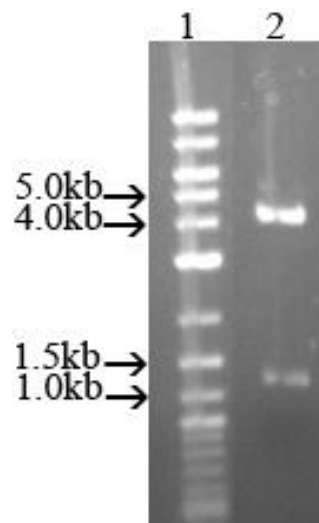
Figure 3.13 The separated DNA fragments of digested plasmid by ethidium bromide stained agarose gel electrophoresis, *the expected fragments given in parenthesis **Lane 1,** 1 kb DNA ladder, **lane 2,** pAKV1 digested with BglII and SpeI (4.1 kb and 1.3 kb).*

3.2.7 Sequencing of *opptktA3* in pMK11

The *opptktA3* gene in pMK11 was sequenced using the PCR-primers described in table 2.4. The pairwise alignment of PCR-opptktA3 and the PCR sequence designed on Clone Manager program showed to be 100% identical (appendix H).

3.2.8 Construction of the *tktA3* gene inactivation vector pMK12

The vector plasmid pHE216 was digested with *SacI* and *NotI* to give two fragments 11.4 kb and 4.1 kb (Figure 3.14a) .The pMK11 plasmid was digested with *SacI* and *PspOMI* to give fragments of 2.6 kb and 3.1 kb (Figure 3.14a). The 11.4 kb vector fragment was ligated to the 2.6 kb insert fragment. The ligation mixture was transformed into *E. coli S17.1 λ pir* competent cells. The transformed cells were selected on LA plate with Amp overnight at 37°C. Plasmids from six transformants were isolated and digested with *EcorV*. The expected fragments (8.7 kb, 4.1 kb and 1.1 kb) observed on the gel (Figure 3.14b shows only one colony). The plasmid chosen for further work of *tktA3* inactive gene was designed pMK12.
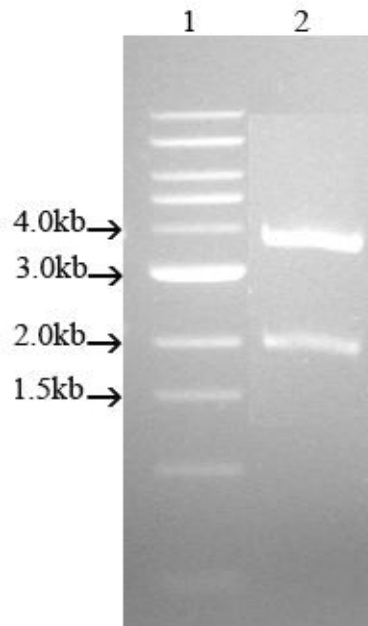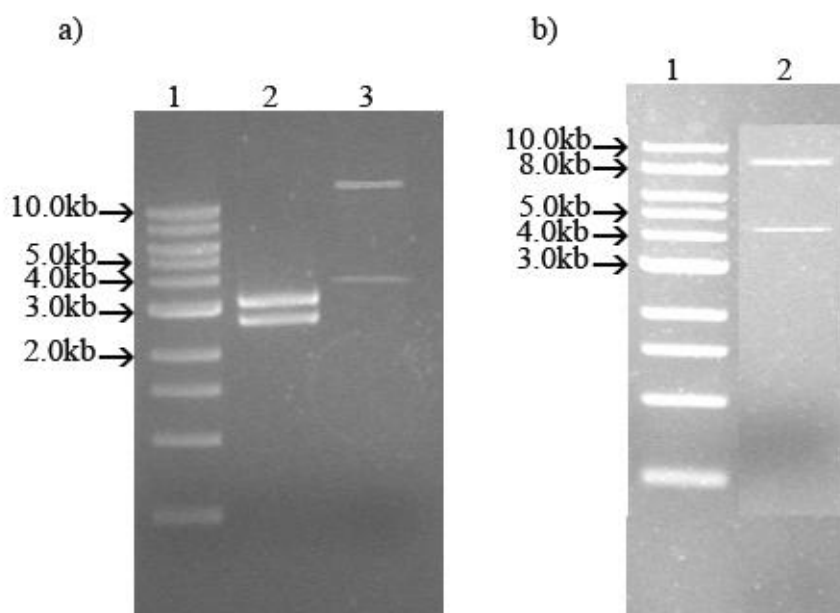
Figure 3.14 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments given in paranthesis **a) Lane 1,** 1kb DNA ladder, **lane 2,** pMK11 digested with SacI and PspOMI (2.6 kb and 3.1kb), **lane 3,** pHE216 digested with SacI* and *NotI (11.4 kb and 4.1 kb)* ***b) Lane 1,*** *1 kb DNA ladder,* ***lane 2,*** *pMK12 controlled by digestion with EcorV (8.7 kb, 4.1 kb and 1.1 kb) the 1.1kb band is not shown due to low image resolution.*

3.3   Gene inactivation

3.3.1 Conjugation of pMK5, pMK9 and pMK12 to *A. vinelandii* ATCC12518

The plasmids containing the deletion mutation (table 2.2) in the *tktA1⁻, tktA2⁻* and *tktA3⁻* (pMK5, pMK9 and pMK12 respectively) genes were transferred to *A. vinelandii* ATCC12518 from *E. coli S17.1 λ pir* by conjugation.
The *E. coli S17.1 λ pir* cells contain the genes responsible for the transfer of genetic material between cells through conjugation. *S17.1 λ pir* also contains the gene *pir* which codes for protein π required during the replication of R6K-ori (for more details see subsection 2.2.9.2).

Conjugants from pMK5, pMK9 and pMK12 to *A.v.ATCC12518* were selected on BA and Sp. No colonies were observed from conjugation of pMK9 containing *opptktA2* to *A.v.ATCC12518*. More than 12 transconjugants colonies of pMK5 and pMK12 to *A.v.ATCC12518* were observed on the selected plate of BA and Sp.

Four colonies from each conjugation of pMK5 and pMK12 to *A.v.ATCC12518* were then selected as described in subsection 2.2.9.2.1 and named as follows; *A. vinelandii tktA1-* k1, *A. vinelandii tktA1-*k2, *A. vinelandii tktA1-* k3, *A. vinelandii tktA1-* k4, *A. vinelandii tktA3-* k1, *A. vinelandii tktA3-*k2, *A. vinelandii tkA3-*k3 and *A. vinelandii tktA3-*k4.

The transconjugants after five times of subculturing in BM medium with antibiotics used for initial selection (Sp) to allow the segregation of the *A. vinelandii* chromosome copies to take place. Followed by subculturing in BM without Sp to allow plasmid excision. The serial dilution of the transconjugantes (pMK5 and pMK12 to *A.v.ATCC12518*) were selected on BA plates with 5% sucrose.

Nine sucrose-resistant colonies from each transconjugant were picked and tested for their spectinomycin sensetivity on BA containing 2% glucose with added Sp and on BA containing 2% glucose without Sp to indicated the lose of suicide vector sequence. None of the selected colonies were able to grow on either of these plates. This indicated that the nine colonies were not *A. vinelandii*.

Previous studies carried out on transketolase mutants in both yeast and *E. coli* that failed to grow without aromatic supplement (Josephson & Fraenkel, 1969, Schaaff-Gerstenschlager & Zimmermann, 1993). The RA1 medium was favoured over BM in cultivation of mutant *A. vinelandii* during the selection procedures. If this was the case for *A. vinelnadii* as well, using a rich medium could aid in obtaining the combinants.

The same procedures used for selecting the recombinants in BM was repeated using RA1 medium. It was observed better growth in RA1 medium than in BM (see discussion). The *A.v.ATCC12518:pMK5* and *A.v.ATCC12518:pMK12* were able to grow on RA1 plates containing 2% glucose with added Sp and without Sp after their selection on RA1 plates with sucrose 5%.

All the recombinants were Sp^R this should not be happened as after a double-crossover event (the vector is excised from the recipient cell), mutants that had a successful gene replacement should be sucrose^R and Sp^S. The mutants also can be reverted to the wild-type (plasmid with mutant version excised from the cell leaving the wild-type version of the gene in the chromosome). However with sucssesful recombination no cells should be both sucrose resistance and Sp^R. Both transconjugants (*A.v.ATCC12518:pMK5* and *A.v.ATCC12518:pMK12*) were Sp^R and sucrose^R.

Based on these observations the genomic DNA was isolated from two transconjugate of *A.v.ATCC12518:pMK12* and controlled by PCR using primers described in table 2.3. The Figure 3.15 shows the PCR product separated on gel electrophoresis with a band size of 2.2 kb, which corresponds the wild-type of *tktA3,* in case the mutant *tktA3*$^-$ was obtained the size had to be 1.7 kb.
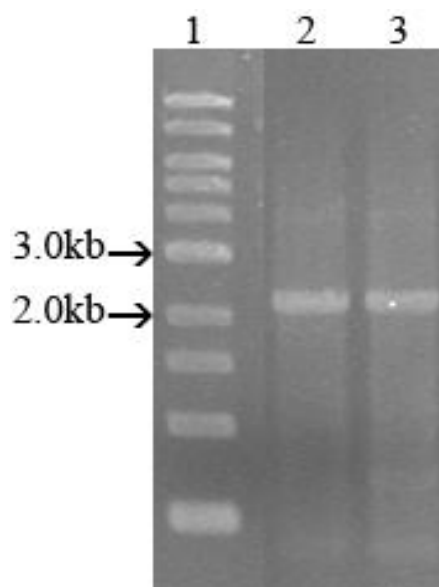
Figure 3.15 The PCR products of two trans-conjugated colonies from A.v.ATCC12518:pMK12. Lane 1, 1 kb DNA ladder, lane 2 and 3, PCR product of 2.2 kb, indicated the wild-type of tktA3 gene.

This unexpected result of obtaining a wild-type transketolase rather than the mutant, led to examine the first obtained colonies selected on BA with Sp after the conjugation of pMK5 and pMK12 to *A.v.ATCC12518* (before the homologous recombination).

The colonies were selected on RA1 plates with 5% sucrose and Sp. These colonies had to contain both the chromosomal DNA of *A. vinelandii* (able to grow on BA or/and RA1) and the plasmids pMK5 or pMK12, which are sucrose sensitive and Sp^R as the both phenotypes are plasmid, related.

The colonies were both sucrose and Sp resistant suggesting that the vector plasmid conjugation to *A. vinelandii* did not occur. The growth of *A. vinelandii* cells on Sp suggesting that different mutational events could give rise to spectinomycin-resistant progeny spontaneously (see discussion). Since no mutant found after several repeated experiments, the strategy of inactivation of these genes in order to elucidate their biological role in *A. vinelandii* had to be given up.

3.4 Over-expression, purification and activity measuring of transketolase recombinant protein

Another part on the study was to try to show that the three genes actually encode transketolases.

The three transketolase genes (*tktA1, tktA2* and *tktA3*) were cloned into the over-expression vector pHE263 under control of the m-toluate inducible *Pm*-promoter.

The expression plasmids pMK13, pMK14 and pMK15 containing *tktA3, tktA1* and *tktA2* respectively were to be transformed into an *E. coli K-12* production strain. The three transketolase genes were to be over-expressed, purified and studied using enzyme assay described in subsection 2.2.11.2.

3.4.1 Construction of three transketolase over-expression plasmids in *E.coli*

Transketolase genes were digested using restriction enzymes from previously constructed vectors pMK2, pMK6 and pAKV1 containing *tktA1, tktA2 and tktA3* respectively. Fragments containing the three genes were obtained by digesting these plasmids with *NotI* and *NdeI*. The fragments of insert obtained were *tktA1* of 2.2 kb, *tktA2* 2.1.kb and *tktA3* 2.2 kb (Figure 3.16). The vector plasmid pHE263 also was digested with *NotI* and *NdeI* to give two fragments of 8.3 kb and 1.2 kb (Figure 3.16).

The vector fragment of 8.3 kb was ligated to *tktA3* gene and transformed to *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. Plasmids from four transfomants were isolation and digested with *NotI* and *NdeI*. The expected fragments (8.3 kb and 2.2 kb) observed on the gel (Figure 3.17 colony no. 3 was used further). The plasmid was designated pMK13.

The *tktA1* gene was ligated to the pHE263 vector fragment of 8.3 kb and transformed into *E. coli S17.1 λ pir* competent cells and selected on LA with Amp. The four obtained colonies were inoculated for plasmid isolation. The isolated plasmids were digested with *NotI* and *NdeI* for control and the expected fragments (8.3 kb and 2.1 kb) observed on the gel (Figure 3.17) colony no.2 was used further. The designated plasmid was pMK14.

The *tktA2* gene was ligated to the 8.3 kb DNA fragment of pHE263 and transformed into *E. coli S17.1* competent cells and selected on LA with Amp. The four obtained colonies were inoculated for plasmid isolation and the isolated plasmids were digested with *NotI* and *NdeI*. The expected fragments (8.3 kb and 2.2 kb) were observed on the gel (Figure 3.17 colony no.1 was used further). The plasmid was designated pMK15.
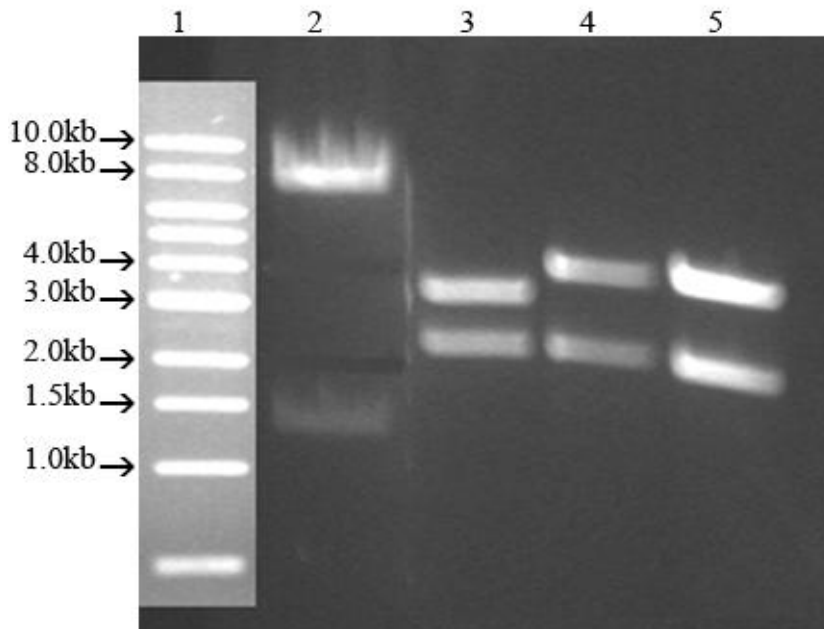
Figure 3.16 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are given in parenthesis.* ***Lane 1,*** *DNA 1kb ladder,* ***lane 2****, pHE263 digested with NotI and NdeI (8.3 kb and 1.2 kb),* ***lane 3,***

*pAKV1 digested with NotI and NdeI (3.2 kb and 2.2 kb), **lane 4**, pMK2 digested with NotI and NdeI (3.9 kb and 2.2 kb), **lane 5**, pMK6 digested with NotI and NdeI (3.9 kb and 2.1 kb).*
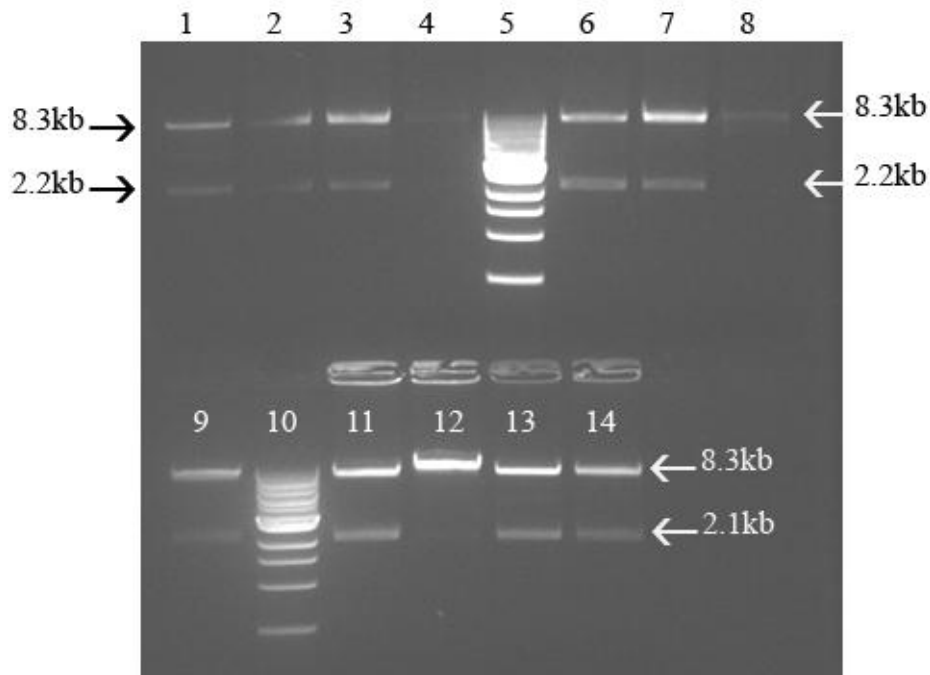
***Figure 3.17 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis,** the expected fragments are given in parenthesis. **Lane 1-4,** pMK13 digested with NotI and NdeI (8.3 kb and 2.2 kb), lane 4, no plasmid obtained, **lane 5,** DNA 1 kb ladder, **lane 6-9,** pMK14 digested with NotI and NdeI (8.3 kb and 2.2 kb), **lane 10,** DNA 1 kb ladder, **lane 11-14,** pMK15 digested with NotI and NdeI (8.3 kb and 2.1 kb) lane 12 didn't show the correct bands expected).*

The three constructed vector plasmids (pMK13, pMK14 and pMK15) were transformed into *E. coli K-12* production strain *RV308*. No growth was observed after the transformation of plasmid DNA from pMK13 plasmid vector. One colony of each of pMK14 and pMK15 was obtained. The isolated plasmid from pMK14 digested with *NotI* and *NdeI* showed no band or plasmid DNA on the gel. While the isolated plasmid from pMK15 digested with the same restriction enzymes showed one band of 8.0 kb size which does not corresponds the expected bands (8.3 kb and 2.1 kb) (Figure 3.18). Thus in this strain no transformants containing any of these three plasmids were observed (see discussion).
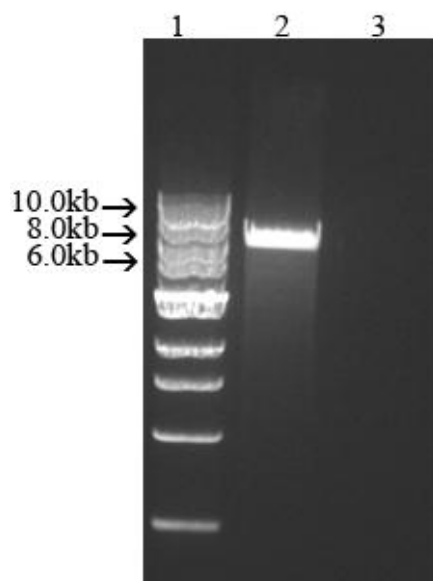
Figure 3.18 The separated DNA fragments of digested plasmids by ethidium bromide stained agarose gel electrophoresis, *the expected fragments are given in parenthesis.* ***Lane 1,*** *DNA 1 kb ladder,* ***lane 2,*** *pMK15 digested with NotI and NdeI (8.0 kb),* ***lane 3,*** *pMK14 digested with NotI and NdeI did not show any band or plasmid DNA.*

The three plasmids were also transformed into *E. coli DH5α*, and no growth of this strain containing any of these plasmids are observed either.
For these reasons the three vector plasmids pMK13, pMK14 and pMK15 transformed originally to *E. coli S17.1 λ pir* and *E. coli S17.1* were used to study transketolase activity.

3.4.2 Substrate specificity and kinetic properties of the recombinant transketolase (enzyme assay)

The strains containing the expression vector plasmids (pMK13, pMK14, pMK15 and pHE263 as negative control), were grown in 3xLB medium with the addition of m-toluic acid (final concentration 0.5 mM) as an inducer for *Pm*-promoter. The cell free extract was prepared as described in subsection 2.2.11.1.

The enzyme assay was carried out as described in subsection 2.2.11.2, and the transketolase activity was determined by measuring spectrophotometrically the increase in NADH absorbance at 340 nm and 37°C with xyloluse-5- phosphate and ribose-5-phosphate as substrates.

The cuvettes containing the assay mixture was incubated for 45 min prior to enzyme addition in order to stabilize the substrates. No protein activity was observed of measuring samples containing transketolase (pMK13, pMK14 and pMK15) comparing to the control sample pHE263. The graphs of the spectrophotometrically activity measurements are shown in appendix I.

3.4.3 Protein expression and size determination of *tktA1*, *tktA2* and *tktA3* genes in *E. coli* using SDS-PAGE

The lack of protein activity could indicate poor expression level of the proteins. Thus the three protein extracts were also analyzed on SDS-PAGE.

The, *tktA1* and *tktA2* genes products are 74.6 kDa, and the *tktA3* gene products is 73.8 kDa, respectively. The protein samples were applied in two concentrations; 10 μl of the cell free extract and 1 μl of the cell free diluted with Tris-HCl buffer (final concentration 50 mM, pH 7.7) to 10 μl. From Figure 3.19 is seen that the molecular mass of the two products of the transketolase genes (*tktA2* and *tktA3*) show presence of bands corresponding the calculated molecular weight. The control extract (pHE263) did not show a band of this range.
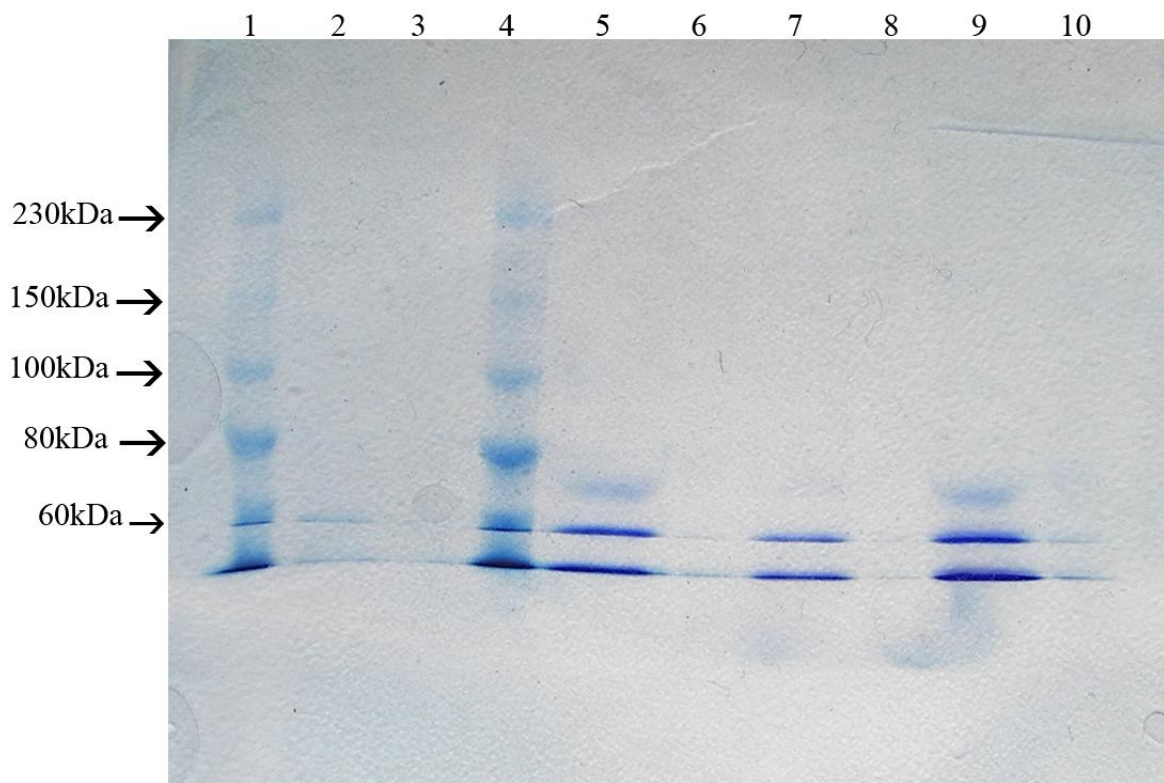
Figure 3.19 Separation of transketolase proteins on 10% SDS-PAGE Coomassie-stained gel shows in well 1, protein standard, well 2 pHE263 the control plasmid (undiluted), well 3,pHE263 diluted, well 4, protein standard, well 5 TktA3 (undiluted) of 73.8 kDa, well 6 Tkt3 (diluted) no protein observed, well 7, Tkt1 (undiluted) no protein observed, well 8, Tkt1 (diluted) no protein observed, well 9, TktA2 (undiluted) of 74.6 kDa, and well 10, TktA2 (diluted) no protein observed.

3.4.4 Purification of recombinant transketolase proteins (TktA1, TktA2 and TktA3) by Column chromatography

The proteins seemed to be expressed the lack of protein activity could be caused by impurities repressors or even competing enzyme activities in the extract, leading to study protein

expression and activity after its purification form mixture using a higher amount of extract in the assay.

The OD_{600} and volume was measured prior to harvesting the cells by centrifugation and sonication. The OD_{600} and volume of pHE263 (negative control) was 0.444 and 230 ml, of pMK13 (*tktA3*) was 0.564 and 230 ml, pMK14 (*tktA1*) was 0.500 and 240 ml and pMK15 (*tktA2*) was 0.930 and 235 ml. The three transketolase proteins were purified by FPLC. The anionic exchange chromatography was carried out using a buffer with increased salt concentration.
The fractions obtained were analysed using SDS-PAGE to see which fraction contained the protein of interest. The chromatogram graphs for each protein sample are shown in the Figures 3.20, 3.21 and 3.22.

The three Figures 3.20, 3.21 and 3.22 show the chromatogram for each of the three transketolase proteins separated from mixture. The fractions of observed picks for each protein sample on the chromatogram were separated on SDS-PAGE (Figure 3.23 and 3.24).
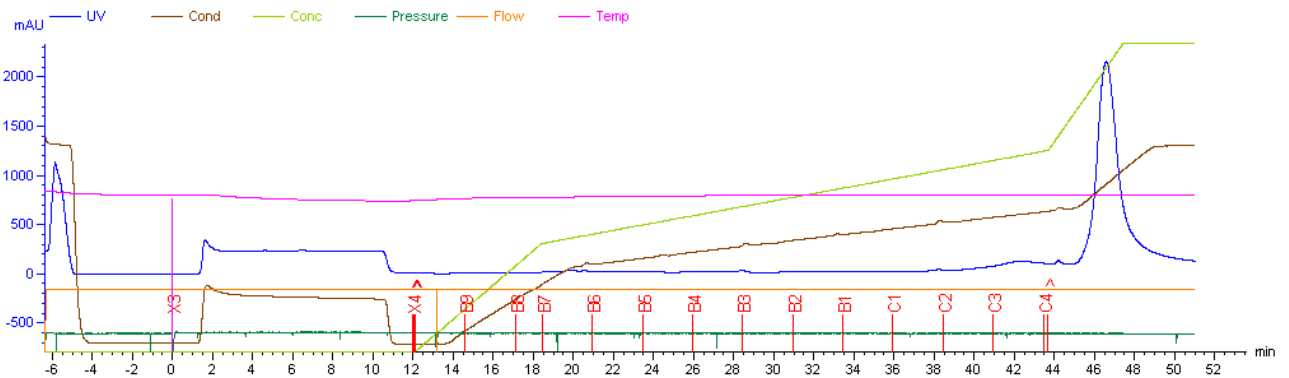
Figure 3.20 The chromatogram of FPLC purification of the cell free extract from S 17.1 λ pir (pMK14) vector plasmid.
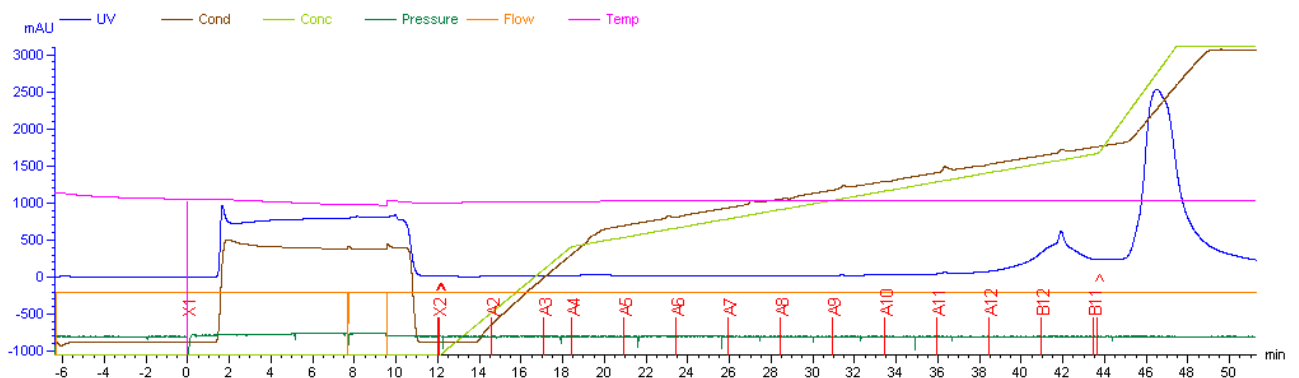
Figure 3.21 The chromatogram of FPLC purification of the cell free extract from S 17.1 λ pir (pMK13) vector plasmid.
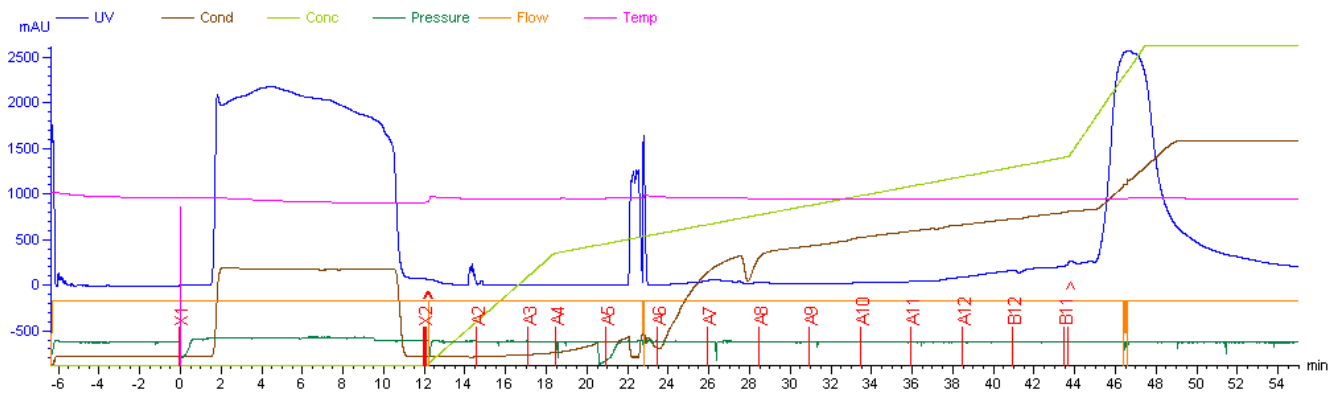
Figure 3.22 The chromatogram of FPLC purification of the cell free extract from S 17.1 λ pir (pMK15) vector plasmid.

The band corresponding the TktA1 protien applied from the separated fractions x3, B7, B6, B4, B3, C1, C2, C3 and C4 (Figure 3.20) was not observed on SDS-PAGE (Figure 3.23 and 3.24). The TktA2 protein was in fraction x1 (Figure 3.22) and is observed on the SDS-PAGE as a band of 74.6 kDa (Figure 3.24). The TktA3 is observed on the SDS-PAGE as a 73.8 kDa protein band (Figure 3.23, well 5). This corresponds to the x1 fraction in Figure 3.21.
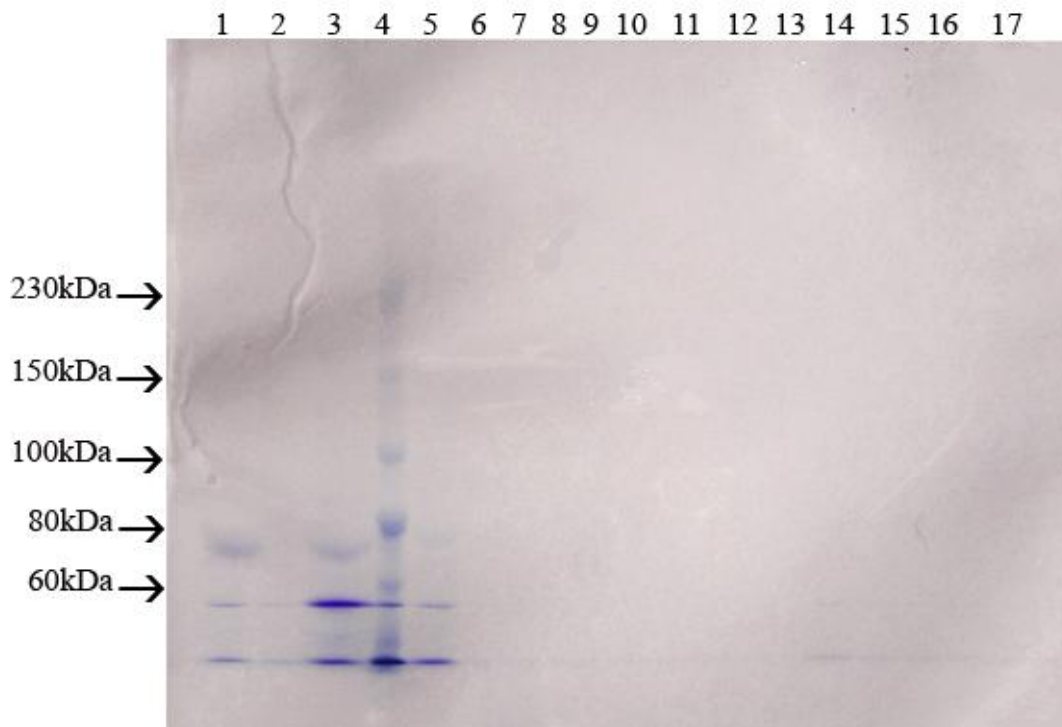
Figure 3.23 The separated fractions obtained from the chromatography on 10% SDS-PAGE Coomassie-stained gel shows in well 1, *cell free extract of tktA3 shows a band at 73.8 kDa (due to low picture resolution the band weekly observed on the printed figure),***well 2,** *cell free extract of tktA1, no band observed,* **well 3,** *cell free extract of tktA2 at 74.6 kDa,* **well 4, 230 kDa protein ladder,* **well 5,** *fraction x1 of tktA3 shows the separated protein of 74.6 kDa,* **well 6-12,** *fractions A4, A9, A10, A11, A12, B12, B11 obtained from cell free extract of tkt3 (showed no bands),* **well 13-17,** *fraction obtained from tktA1, x3, B7, B6, B5,and B4 no proteins were observed among these fractions.*
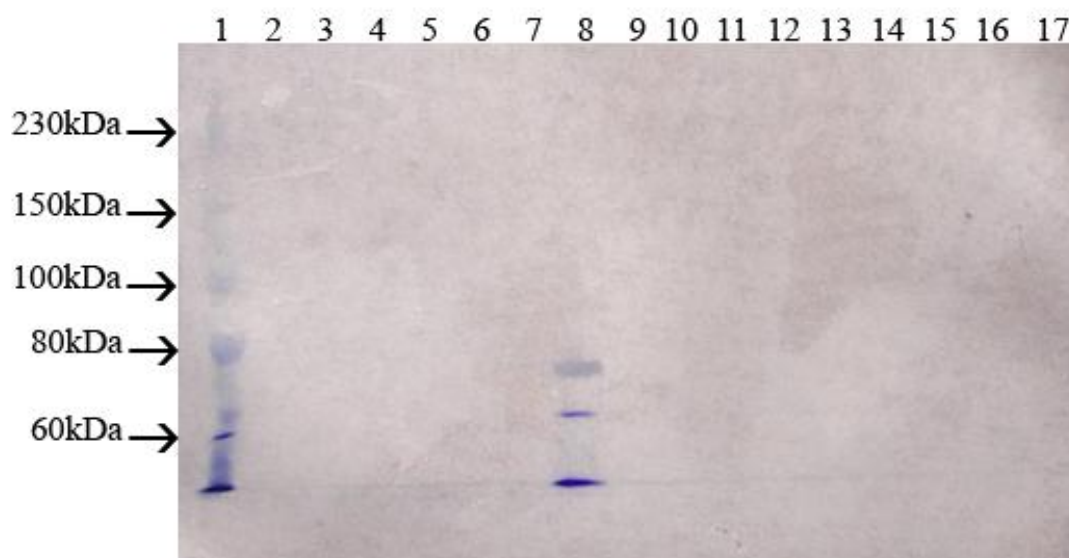
Figure 3.24 The separated fraction obtained from the chromatography on 10% SDS-PAGE Coomassie-stained gel shows in well 1 230 kDa protein ladder, well 2-7 fractions obtained from the separation of tktA1 gene product B3, C1, C2, C3 and C4 (no bands observed), well 8 fraction x1 of tktA2 gene product of 73.8 kDa, well 9-17 other fractions A2, A3, A5, A6, A7, A8, A12, B12 and B11 obtained from cell free extract of tktA2 (showed no bands on the gel).

The transketolase activity was measured for the second time, using the fractions x1 of TktA2 (Figure 3.22) and fraction x1 of TktA3 (Figure 3.21) purified by column chromatography. No protein activity was observed (see discussion).

3.4.5 Measuring the protein concentration of TktA2 and TktA3

The concentration of TktA2 and TktA3 proteins obtained from FPLC and the free cell extract was measured using Bio-Rad Protein Assay. BSA was used as standard protein. The absorbanse was measured at room temperature at 595 nm. The data and the standard curve are shown in Appendix F.

Table 3.1 describes the concentration calculate from the equation of linearity for the standard curve and the concentration of protein in sample.

Table 3.1 The concentration of the TktA2 and TktA3 proteins.

Proteins	*Concentration calculated from the standard curve (µg/ml)*	*Concentration calculated in sample (mg/ml)*
Protein amount in Tkt A2 cell free extract	1.8	0.360
Protein amount in TktA2 x1 fraction obtained from FPLC	5.0	1.0
Protein amount in Tkt A3 cell free extract	6.3	1.260

Protein amount in TktA3 x1 fraction obtained from FPLC	3.5	0.700

3.5 Bioinfomatical study of transketolase genes *(tktA1, tkA2 and tktA3)* from *A. vinelandii* and their organization in the genome

The microorganism of interest in this work is *A. vinelandii*, which has been sequenced and showed to have three transketolase copies *tktA1*, *tktA2* and *tktA3* (Setubal et al., 2009). These three gene copies are located in the chromosome, with a total length of *tktA1* is 2.091 bp (encodes transketolase protein of 672 kDa), *tktA2* is 2.019 bp (protein of 672 kDa) and *tktA3* is 1.998 bp (protein of 665 kDa).

where *tktA1* and *tktA2* are most identical (98% similarity) with four base pairs difference at 975 bp (A=G), 1125 bp (G=A), 1290 bp (C=T) and 1878 bp (G=T) locations in the genes. The *tktA3* is smaller and less identical to the previous two (75% similarity). A multiple alignment of the three genes is shown in appendix G.
Theoretically the close distances between genes tend to be recognized as genes localized in the same operon. Distances of more than 50 bp tend to be recognized as genes in two different operons. The prokaryotic promoter is located -10 to -35 from the transcription start, Shine-Dalgarno sequence (SD) is -8 bp to -12 bp upstream from the start codon (ATG/GTG) (Figure 3.25). Still, no absolute rules for the distances between co-transcribed genes.
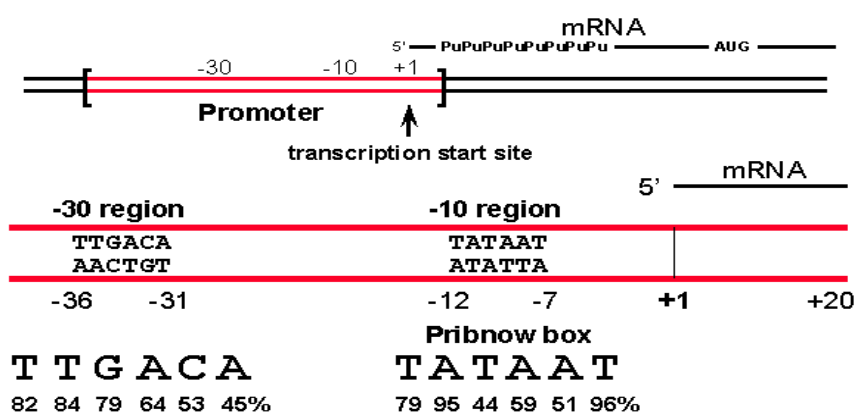
Figure 3.25 The promoter and the transcription start site in prokaryotes (Watson et al., 2004).

The distance in base pairs from the stop codon of the transketolase gene *tktA1* and the start of the down stream gene transaldolase B1 (*talB1*) 9 bp (Figure 3.26). The *talB1* gene can be proposed and considered to be located in the same operon with *tktA1* as the space between the two genes is to close. The *avin_ 22130* is located downstream of the *talB1* gene and overlap with it in 16 kb. *rpiB* has a distance from its start codon and the stop codon of *rpiB* is 274 bp

upstream of gene (*avin_22130*) in 274 kb. The genes *tktA1*, *talB1* and *avin_22130* can be proposed to be located in the same operon.

The ribose 5-phosphate isomerase gene (*rpiA1*) is, located 307 bp down stream of *tktA1* gene and ranges from 2,215.359 kb to 2,216.033 kb. The difference in kilo base from the start codon of *tktA1* and the stop codon of *rpiA1* is 0.307 kb (307 bp). The *rpiA1* is located far enough from the *tktA1*, that is probably not located in the same operon (Figure 3.26).

Figure 3.26 The location of transketolase A1 (tktA-1) gene in the genome of A. vinelandii shown from 2210637-2217694 kb. *"rpiB" indicates for ribose 5 phosphate isomerase B gene," Avin22130"HAD-superfamily hydrolase, "talB1"transaldolase B1, "tktA-1"transketolase A1,"rpiA-1"ribose 5 phosphate isomerase, "Avin_22170" sugar-binding domain-containing protein and "Avin_22180" hypothetical protein.*

tktA2 is located in the genome 9 bp down stream gene is transaldolase B2 (*talB2*) which starts at 2,808 792 kb. Downstream of *talB2* is *avin_27350*, which is started at 2,809744 kb overlap with the *talB2* in 16 kb, and the next downsteraming gene is *mtID* located 18 kb far from *avin_27350*. *mtlD* is the last gene in a probable operon containg *mtlK*, *mtlG*, *mtlF* and *mtlE*. These corresponding gene products are all involved in.

The gene located upstream to *tktA2* is xylulokinase (*mtlY*) ends at 2,806.762 kb with a difference of 3 kb (3000 bp) from the start codon of the *tktA2*.

The *talB2* and *avin_27350* are probably located in the same operon, the other genes appears to be closely located (Figure 3.27).
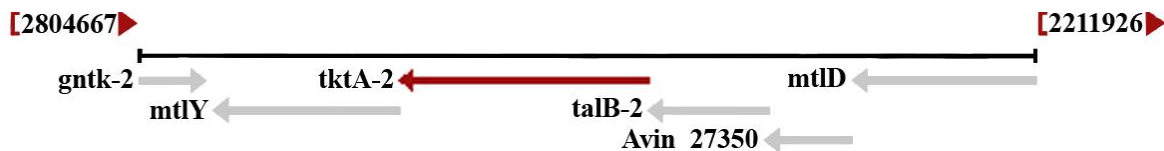
Figure 3.27 The location of transketolase-A2 gene in the genome of A. vinelandii shown from 2804667 -2811926 kb. *"mtlD" manitol dehydrogenase, "Avin_27350" HAD-superfamily hydrolase, "talB-2" transaldolase B2,"tktA-2" transketolase A2,"mtlY" xylokinase and "gntK-2" gluconokinase.*

The *tktA3* gene starting from 1,535.097 kb to 1,535.100 kb (Figure 3.28), the gene located upstream is erythrose 4 phosphate dehydrogenase (*epd*) which ends at 1,537.218 kb with a difference between the to in 0.121kb (121bp). The hypothetical protein gene *Avin_05550* located down stream shows to overlap with *tktA3* gene. As *Avin_05550* starts at location of 1,535.116 kb in the genome and the *tktA3* ends at 1,535.100 kb they overlap in 16 kb (160 bp). The *avin_05550* codes for a small hypotherical protein of (162 aa). Not all ORFs are genes, and the real genes have to contain a promoter for example in order to be transcribed. The *avin_05550* can be a non coding DNA fragment found in the genome. Overlapping genes can be considered as two adjacent genes (genes in the same strand) where their coding regions are overlapping. In such cases bioinformatic studies are not enough to consider the

overlapping as two genes and have to be tested experimentally. When two independent genes overlap, two separate products have to be identified.
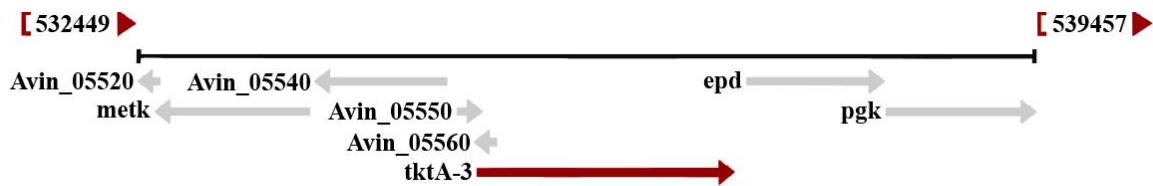
Figure 3.28 The location of transketolase-A3 gene in the genome of A. vinelandii shown from 532449-539457 k. *"Avin_05520" hypothetical protein, "metK" S-adenosylmethionine synthetase,"Avin_05540" ArsR family regulatory protein, "Avin_05550"hypothetical protein, "Avin_05560" hypothetical protein,"tktA-3" transketolase A3, "epd" erythrose 4 phosphate dehydrogenase and "pgk" phosphoglycerate kinase.*

The *talB1* (transaldolase) and *rpiB* (ribose 5-phosphate isomerase) gene are found in the same proposed operon with *tktA1* and participate in the PPP, the closely located *rpi-A1* gene also participates in the PPP.
The *tktA2* gene is located close to *talB2* gene which is also participates in the PPP. The *avin_27350* is probably located in the same operon with *talB2*.
Avin_05550 is overlapping with *tktA3,* This small DNA fragment that codes for hypothetical protein following NCBI, has a low probability to be a real gene. *epd* is closely located to *tktA3* and is responcible for the conversion of 4-phosphoerytronate to erythrose-4-phosphate in the PPP. The *pgk* located downstream of *epd* converts 1,3-bisphosphoglycerate to glycerate 3-phosphate also involved in the central carbohydrate metabolism (Figure 3.29).

Both TktA1 and TktA3 proteins participate in the pentose phosphate pathway, metabolic pathway and biosynthesis of secondary metabolites. And *tktA2* gene can be proposed to participate in the PPP due to its location in the chromosome with other genes (*talB-2*) that participated in PP pathway. Thus the three transketolase genes and the genes located either in the same operon or in close distance participates in the PPP and other central carbohydrate metabolism pathways (following KEGG) (Figure 3.29). The transketolase being coupled to several different pathways from which the sugars can be metabolised. This is because the bacteria tend to use different pathways to obtained the target sugar under different conditions. Not all the neighboring genes to the three transketolase are parts of the PPP or the central carbohydrate metabolism connected to it.
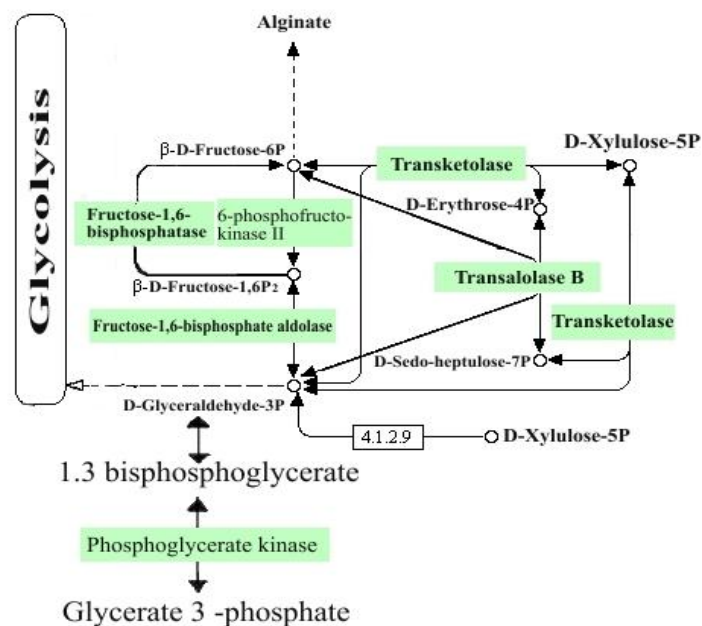
Figure 3.29 Transketolase and the closely located genes involved in central carbon metabolism.

Transketolase genes from *A. vinelandii* were blasted on NCBI and showed to be similar to several strains of *Pseudomonades aeroginosa, tktA1* was 78% identical, score 1024 and E-value 0.0, *tktA2* was 78% similar, score 1024 and E-value 0.0 and *tktA3* was 86% identical, score 2023 and E-value 0.0. *P. aeroginosa* has only one transketolase which gave higher similarity against *tktA3* of *A. vinelandii.* The mostly studied transketolase in prokaryotes is *E. coli tktA* and *tktB*, a multiple alginment of the three transketolase genes to both *E. coli* transketolase is shown in appendix G. *tktB* of *E. coli* appears to be the most similar to the three *A. vinelandii* transketolase with 66-68% similarity (appendix G).

The conserved domains for the three transketolase gene products were analyzed using pfam, Prosite and InterPro and are shown in the tables 3.2, 3.3 and 3.4.

Table 3.2 The concerved domains in the protein of *tktA1* gene.

Domains	E-value	source
Transketolase-N	$9e^{-155}$	pfam
Transket-pyr	$1.6e^{-43}$	pfam
Transketolase-C	$9.6e^{-07}$	pfam

Table 3.3 The concerved domains in the protein of *tktA2* gene.

Domains	E-value	source
Transketolase-N	$9e^{-155}$	pfam
Transket-pyr	$1.6e^{-43}$	pfam
Transketolase-C	$9.6e^{-07}$	pfam

Table 3.4 The concerved domains in the protein of *tktA3* gene.

Domains	E-value	source
Transketolase-N	$4.9\mathrm{e}^{-159}$	pfam
Transket-pyr	$9.4\mathrm{e}^{-45}$	pfam
Transketolase-C	$1\mathrm{e}^{-09}$	pfam

The three transketolase proteins appear to have the same domains transketolase-N domain, transketolase-pyr and the transketolase-c domain.

4 Discussion

The first part of this work was to study the impact of the in frame deletion mutants of three transketolase gene (*tktA1, tktA2* and *tktA3*) on alginate production as the transketolase participates in the non oxidative branch of the pentose phosphate pathway to produce fructose-6 phosphate, the first precursor in alginate biosynthesis. It would also have been intrestingly to estimate the influence of these three genes on cyst formation in *A. vinelandii* due to the importance of alginate in this process by comparing the level of alginate production in the mutants to the wild-type.

4.1   Three genes of transketolase in *A. vinelandii*

The *A. vinelandii* has three transketolase gene *tktA1*, *tktA2* and *tktA3*, all believed to be part of the PPP and other pathways related to it. The nucleotide sequences of both *tktA1* and *tktA2* were almost identical to each other while they differ from *tktA3* by approximately 15%. Having three genes, can be explained out the nutrients availability for the bacteria. Under conditions were there are enough food, the bacteria may need to turn only one transketolase gene ON, keeping the others OFF or silent. As the nutrients in the medium become less available the bacteria then needs to turn all the transketolase genes ON, in order to obtain enough nutrients/energy through different pathways to stay alive. It could be also that the higher growth of bacteria needs more than one transketolase.
A. vinelandii has an important process, the cyst formation that needs alginate. Keeping in mind that *A. vinelandii* produces alginate also in it's vegetative stage. The alginate synthesized through transketolase, as the bacteria need to form cyst it could be more efficient of having three genes that connects several pathways in order to obtain enough alginate for this process.
Here the obvious question rises of why so many gene copies are needed by this organism. In addition to three transketolase, the bacteria also has seven epimerase genes, at least six lyase and three transaldolase and as well several other genes that are found in more than one copy. All genes stated are involved in alginate synthesis or modification (degradation) and can be proposed to be important in the cyst formation process.

4.2   Identification of transketolase mutants

In many cases, it is essential to make deletion mutation rather than antibiotic-resistance allelic/gene replacement because the antibiotic-resistance replacement can lead to disruption of expression of genes downstream of the inserted antibiotics-resistance marker. Thus in this work three in frame deletions of 1200 bp in the three transketolase genes were constructed using PCR primers, in order to disrupt the transketolase genes only and not the others found in the same operon. The three cloned DNA fragments were verified by DNA sequencing and showed to be 100% identical to those designated using Clone Manager.

The transformants of the resulting plasmids into *A. vinelandii* ATCC12518 were selected on BA and Sp and showed the growth of only of *A.v.ATCC12518:pMK5* and *A.v. ATCC12518:pMK12*. No growth was observed of *A.v.ATCC12518:pMK9*.

74

However even after several experiments using both Burks agar and RA1, no deletion mutants were observed. As no mutant transketolase was isolated in *A. vinelandii*, the speculations of better growth with aromatic supplement (RA1) as are the case for yeast and *E. coli* transketolase mutants is not the reason.

It was believed that the high chromosome copy numbers, leads to difficulty in segregation of mutant phenotypes. However, several successful studies carried out on *A. vinelandii* showed that generating mutation within its chromosome is possible either by transposon mutagenesis or by homologous gene replacement (antibiotic resistant cassette or deletion mutations). This experimental strategy of conjugation and selection on *sacB* gene were previously used and showed to be efficient in *A. vinelandii* (Gimmestad et al., 2009).

As stated previously that the three PCR products of the transketolase genes (*opptktA1, opptktA2* and *opptktA3*) were sequenced and showed to be 100% identical to the comparing sequences in the genome. The pHE216 vector has been successfully used in other studies (Ertesvåg, not published). The only difference here is the use of transketolase genes.

Thus, all the arguments stated above are not the reason for not obtaining a null mutants of any of the three transketolases in *A. vinelandii*. Even suggesting that these genes could be essential to *A. vinelandii* based on the observations (i) the wild-type reversion after the selection procedures that gave rise to sucrose^R colonies (vector free).
Or (ii) the vector plasmid carrying the genes died and was not able to survive within the *A. vinelandii* that gave rise to sucrose^R colonies before the recombination, where the colonies at that moment are *A. vinelandii* cells with the plasmid of interest (containing *sacB* and Sp^R genes).
This suggestion cannot be applied based on thinking that the *A. vinelandii* contains high chromosome copy number. The mutation in one of the three genes can be managed by the bacteria using the two others and obtaining the nutrients needed through alternative pathway.

Turning to another explanation for the observed results of the unsuccessful conjugation of the three vectors pMK5, pMK9 and pMK12 to *A. vinelandii* ATCC12518, this might be explained out from that, the parental vector pHE216 showed several unsuccessful cloning experiments of the three genes (*opptktA1, opptktA2* and *opptktA3*) into it and its derivate pMK12. After several repeated transformation experiments it was obtained only one colony containing the pMK5 plasmid, one with pMK9 and six small colonies of PMK12. This also was observed by Vie, A. K. in 2008 as *tktA* (*tktA3*) faced problems during the cloning and transformation to both *E. coli* and *A. vinelandii* specially its low growth. The low transformation frequency to *E. coli* could be the same reason of unsuccessful gene transfer to *A. vinelandii* as well. These observations of low growth and transformation frequency can suggest that the mutant transketolase can be toxic and harmful to both *E. coli* and *A. vinelandii* and its high or mutated production affecting the number of grown colonies.

Finally the Sp^R phenotype noted for both *a.v.ATCC 12518:pMK5* and *a.v.ATCC 12518:pMK12* can be due to mutation occurred that was able to give rise to the Sp^R phenotype spontaneously.

4.3   Over-expression of three transketolase proteins from *A.vinelandii* in *E. coli*

The second part of this work focused on studying *in vitro* the gene products of the three transketolase genes from *A. vinelandii.* The study was carried by identification, characterization and over- expression in *E. coli*.

The three genes were successfully cloned into pHE263 overexpression vector, and were transformed into the *E. coli S17.1 λ pir* cells and sequenced. The nucleotide sequences of the PCR products and the sequences of *A. vinelandii* were 100% identical. Thus, no mutations were introduced during the PCR.
The *E. coli RV308* cells are used and showed to be good production strains for several recombinant proteins in our laboratory (Sletta *et al.*, 2004). Thus, they were chosen to perform this work.
No successful transformation of the plasmid DNA was achieved in *RV308* productive cells. Some colonies in *RV308* were obtained with pMK14 and pMK15, but the isolated plasmids showed wrong band sizes. No transformants were obtained using *E. coli DH5α* either. This observation can suggest that the products of these genes were toxic to both *E. coli DH5α* and *RV308*, thus it was unsuccessful to obtain the correct clones.

The *E. coli S17.1 λ pir* and *E. coli S17.1* strains containing the expression plasmids were grown and the activity of their produced proteins was measured using xylulose 5-phosphate and ribose 5-phosphate as substrates. However no activity was found.
The missing activity could be due to unsuitable assay conditions, low protein expression, repressor impurities or competing enzyme activity in the extract.

Very low growth was observed for all three expression vectors in *E. coli*. The measured OD_{600} was below 2. The product of *tktA1* gene did not show any band, on the Coomassie-stained gel of its purified cell free extract. No band was observed either after the purification of the enzyme from the mixture using the FPLC.
The lack of TktA1 expression can be proposed from its toxicity to the cell, as a result no intracellular accumulation occurs within the host cell and the protease activity in such cases might elevate. The proteases serve as a removal for abnormal, toxic, mis-folded or critical regulatory proteins from accumulation in the cell. Whether the enzyme was proteolytically processed it have to be determined experimentally through finding the peptide sequence typical for the secretory protein, in addition comparing the N-terminal sequence of the extracellular protein to its predicted amino acid sequence.
It can also be suggested that the Tkt1 mis-fold and thus cannot accumulate in the cell, or that the *E. coli* cells lack the native regulatory systems for transcription of *tktA1* from *A. vinelandii*.
Finally TktA1 differ from TktA2 by four base pairs where the codon usage can be suggested as a reason of unsuccessful expression of this protein in the *E. coli* cells. Even though *E. coli* shows to use nonrandom synonymous codons, some genes contain rarely used codons in *E. coli*. The four different bases do not appear to be a part of these rare codons, and this cannot be the reason for the unsuccessful expression.

Even though all the three transketolase genes were cloned into *E. coli*, only transcription of two transketolases (*tktA2* and *tktA3*) under control of *Pm*-promoter gave rise to recombinant

proteins. The TktA2 and TktA3 were expressed at levels in *E. coli* that were too low for them to be well detected on a Coomassie-stained gel.

The intensity of bands will usually be proportional to the amount of protein where the expression level of the three genes cannot be estimated as 100% due to low band density observed. Both TktA3 and TktA2 showed week bands on the Coomassie-stained gel with the expected sizes (Figure 3.19). Thus, the purity of the enzymes and their specific activities on the substrates could not be determined.

The result of SDS-PAGE demonstrated the presence of the *tktA2* and *tktA3* gene products and the missing activity can be suggested as; the proteins were secreted to the supernatant but might be nonfunctional in this environment thus lack the enzyme activity, the enzyme activity was not detected also due to the low growth of these proteins even though they showed bands on the SDS-PAGE. The assay used was unsuitable, because it could be detected some activity as the transketolase already presents in *E. coli*.

E. coli cells remains one of the most used organisms for protein expression, although there is no guarantee that a recombinant gene product will accumulate at high level or being biologically active. Thus, many proteins cannot be expressed in *E. coli*, even with strong induced promoter as the *Pm* promoter. In addition, it is stated that the limiting factor for successful expression in *E. coli* is decreasing with high molecular weigh of protein above around 60 kDa (Graslund *et al.*, 2008). The three transketolase is above 70 kDa.

All three transketolase generally can be considered as toxic when produced in the cell. This was observed also during the conjugation of the deletion mutants of these genes into both *E. coli* and *A. vinelandii* concerning the low growth of the transformants. They can also be considered as low growing under the conditions applied during the study. To optimize the expression of these three genes an alternative expression system has to be applied (see chapter 4.4).

4.4   Future perspectives

The three transketolase genes were mutated by in frame deletion, in order to study their influence on alginate and cyst formation. This way of mutation supposedly will leave expression of the down stream genes at normal level. This applied method did not show a significant result and thus another way of mutating these genes can be applied for further studies.

The transketolase genes can be studied by constructing an antibiotics-resistance cassette. The antibiotics resistance cassette can be inserted into the gene, turning it off. During the cloning procedures appearance of any transcription terminators have to be avoided down stream the antibiotics-resistance cassette. It is possible that the transcription of the genes located downstream will be affected. This have to be tested experimentally by comparing the mRNA of both the wild type and the mutants to insure that of the downstream genes in the operon are transcribed. A promoter can also be inserted upstream of the down stream genes in the operon. Mutation located near the transcription start may affect the expression of the gene, thus the SD sequence have to estimated bioinformatically to the genes downstream due to its importance in translational efficiency.

The expression of *tktA1* is believed to be toxic or detrimental to host cell. Thus its transcription and overexpression, has to be tightly regulated under a repressible promoter. Thus, the best system for the bacterial production of toxic proteins completely repressed until

induced. The modified 5'-UTR region of *Pm*-promoter (down regulation) could be useful as a tool to control expression at low level (Lale, 2009). The promoter regulates the expression (no-expression) until the inducer is added. The growth of culture have to me measured at different time intervals and the addition of the inducer can be applied in different concentration in order to find the optimal condition for protein expression.

The assay is believed to be unsuitable. *E. coli* has two transketolase *tktA* and *tktB*, and thus even a minimal transketolase activity had to be observed. The negative results in the assay indicate that it did not work. The transketolase could be obtained commercially and the assay can be tested.

5 Conclusion and suggestions for the further work

Transketolases has been studied since the 1950's in different organisms. In this work the transketolases from *A. vinelandii* was to be studied in order to find their influence of the alginate production and thereby the cyst formation.

To study the influence of three transketolase genes (*tktA1*, *tktA2* and *tktA3*) on alginate production, the in frame deletion mutation was constructed. The three genes had to be inactivated in *A. vinelandii* through a double-crossover event and selected using *sacB* and spectinomycin as markers. Unfortunately no null mutants were obtained. The transformation of these three genes to *E. coli* (the intermediate host) was at low frequency and showed a low growth. The conclusion can be made out of these observations that these genes have a toxic effect to the host cells thus lowering or inhibiting their growth due to toxic accumulation. However more experiments needed to verify this conclusion. A new experimental strategy as antibiotics –resistance cassette can be also tested in order to study these genes further.

The second part of the study was to characterize the three *A. vinelandii* wild-type transketolase genes in *E. coli*. The transformants containing the three plasmids grew at low rate. No activity of the three gene products was observed. TktA2 and TktA3 were expressed at low level and showed week bands on the SDS-PAGE. No band was observed for TktA1. It can be concluded that the missed activity was due to low gene expression or unsuitable assay conditions. The TktA1 was not expressed, did not show any band on the SDS-PAGE and had no measured enzyme activity. This gene product can be considered as toxic to the host cell. All the three genes due to their low expression can be toxic to the cell and need a suitable expression system to increase their production. It would probably be preferable to use a system with close to no expression prior to induction and to induce relatively late in the expression phase.

References

Amersham, (2003) Ion Exchange Chromatography. UK, Amersham biosciences.

Aucamp, J. P., R. J. Martinez-Torres, E. G. Hibbert & P. A. Dalby, (2008) A microplate-based evaluation of complex denaturation pathways: structural stability of *Escherichia coli* transketolase. *Biotechnol Bioeng* **99**: 1303-1310.

Bakkevig, K., H. Sletta, M. Gimmestad, R. Aune, H. Ertesvåg, K. Degnes, B. E. Christensen, T. E. Ellingsen & S. Valla, (2005) Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *J Bacteriol* **187**: 8375-8384.

Ballesteros, F., J. Gonzalez-Lopez, T. de la Rubia, J. Moreno, J. Aneiros & A. Ramos-Cormenzana, (1986a) Growth of *Azotobacter vinelandii* in dialysed soil medium: studies upon the life cycle. *Ann Inst Pasteur Microbiol* **137A**: 55-64.

Ballesteros, F., J. Gonzalezlopez, T. Delarubia, M. V. M. Toledo & A. Ramoscormenzana, (1986b) Some biological characteristics of *Azotobacter vinelandii* (filterable cells) grown in dialyzed soil media. *Microbios* **46**: 159-164.

Bernacchia, G., G. Schwall, F. Lottspeich, F. Salamini & D. Bartels, (1995) The transketolase gene family of the resurrection plant *Craterostigma plantagineum*: differential expression during the rehydration phase. *EMBO J* **14**: 610-618.

Bio-Rad, (1994) Bio-Rad Protein Assay. Source: BioRad Life Science Research. available at http://cores.montana.edu/uploads/Proteomics%20Core/BioradBradford.pdf Downloaded 18.10.2010.

Bjerkan, T. M., B. E. Lillehov, W. I. Strand, G. Skjåk-Bræk, S. Valla & H. Ertesvåg, (2004) Construction and analyses of hybrid *Azotobacter vinelandii* mannuronan C-5 epimerases with new epimerization pattern characteristics. *Biochem J* **381**: 813-821.

Blomfield, I. C., V. Vaughn, R. F. Rest & B. I. Eisenstein, (1991) Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol* **5**: 1447-1457.

Bobest, C. E., (2004) Biochemical and molecular studies of transketolase from *Rhodobacter sphaeroides* and its inactivation by oxygen. USA, The Ohio State University.

Brautaset T., L. R., Valla S., (2009) Positively regulated bacterial expression system. *Microbial biotechnol* **2(1)**: 15-30.

Brownlee, I. A., J. C. Seal, W. Metthew, W. Peter, P. Dettmar & P. J. Pearson, (2009) Application of alginate in food. In: B.H.A. Rehm (ed) Microbiology monographs "Alginate:biology and applications" 13, Germany, Spinger-Verlag.

Campos, M., J. M. Martinez-Salazar, L. Lloret, S. Moreno, C. Nŭnéz, G. Espín & G. Soberon-Chavez, (1996) Characterization of the gene coding for GDP-mannose dehydrogenase (*algD*) from *Azotobacter vinelandii*. *J Bacteriol* **178**: 1793-1799.

Clark, D. P. & N. J. Pazdernik, (2009) Biotechnology : applying the genetic revolution. Amsterdam, Elsevier.

Clementi, F., (1997) Alginate production by *Azotobacter vinelandii*. *Crit Rev Biotechnol* **17**: 327-361.

Cristianini, N. & M. W. Hahn, (2007) Introduction to computational genomics : a case studies approach. Cambridge, Cambridge University Press.

Dale, J., (2002) From gene to genome: concepts and application of DNA technology. UK , John Wiley & Sons Ltd.

Davies, J. C., (2002) *Pseudomonas aeruginosa* in cystic fibrosis: pathogenesis and persistence. *Paediatr Respir Rev* **3**: 128-134.

de Lorenzo, V., I. Cases, M. Herrero & K. N. Timmis, (1993) Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in *Pseudomonas putida* with lacZ-tet bicistronic reporters. *J Bacteriol* **175**: 6902-6907.

Donati, I. & S. Paoletti, (2009) Material Properties of Alginates. In: B.H.A. Rehm (ed). Microbiology monographs "Alginates:Biology and applications" 13. Berlin, Springer-verlag.

Draget, K. I. & T. Cathrine, (2011) Chemical, physical and biologucal properties of alginate and their biomedical implications. *Food Hydrocolloids* **25**: 251-256.

Eidels, L. & M. J. Osborn, (1971) Lipopolysaccharide and aldoheptose biosynthesis in transketolase mutants of *Salmonella typhimurium*. *Proc Natl Acad Sci U S A* **68**: 1673-1677.

Ertesvåg, H., B. Doseth, B. Larsen, G. Skjåk-Bræk & S. Valla, (1994) Cloning and expression of an *Azotobacter vinelandii* mannuronan C-5-epimerase gene. *J Bacteriol* **176**: 2846-2853.

Ertesvåg, H., H. K. Hoidal, H. Schjerven, B. I. Svanem & S. Valla, (1999) Mannuronan C-5-epimerases and their application for *in vitro* and *in vivo* design of new alginates useful in biotechnology. *Metab Eng* **1**: 262-269.

Ertesvåg, H. & S. Valla, (1998) Biosynthesis and applications of alginates. *Polym. Degrad. Stabil.* **59**: 85-91.

Esakova, O. A., L. E. Meshalkina & G. A. Kochetov, (2005) Effects of transketolase cofactors on its conformation and stability. *Life Sci* **78**: 8-13.

Fiedler, E., R. Golbik, G. Schneider, K. Tittmann, H. Neef, S. Konig & G. Hubner, (2001) Examination of donor substrate conversion in yeast transketolase. *J Biol Chem* **276**: 16051-16058.

Galindo, E., C. Pĕna, C. Nŭnéz, D. Segura & G. Espín, (2007) Molecular and bioengineering strategies to improve alginate and polydydroxyalkanoate production by *Azotobacter vinelandii*. *Microb Cell Fact* **6**: 7.

Gimmestad, M., H. Ertesvåg, T. M. Heggeset, O. Aarstad, B. I. Svanem & S. Valla, (2009) Characterization of three new *Azotobacter vinelandii* alginate lyases, one of which is involved in cyst germination. *J Bacteriol* **191**: 4845-4853.

Gimmestad, M., H. Sletta, H. Ertesvåg, K. Bakkevig, S. Jain, S. J. Suh, G. Skjåk-Bræk, T. E. Ellingsen, D. E. Ohman & S. Valla, (2003) The *Pseudomonas fluorescens* AlgG protein, but not its mannuronan C-5-epimerase activity, is needed for alginate polymer formation. *J Bacteriol* **185**: 3515-3523.

Gorin, P. A. J. & J. F. T. Spencer, (1966) Exocellular alginic acid from *Azotobacter vinelandii*. *Can. J Chem* **44**: 993-998.

Graslund, S., P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S. H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C. Y. Kim, L. W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schutz, U. Heinemann, S. Yokoyama, K. Bussow & K. C. Gunsalus, (2008) Protein production and purification. *Nat Methods* **5**: 135-146.

Gyamerah, M., (1997) Kinetics of overexpressed transketolase from *Escherichia coli* JM 107/pQR 700 *Enz Microb Technol* **20**: 127-134.

Hay, I. D., Zahid Ur Rehman, A. Ghafoor & B. H. A. Rehm, (2010) Bacterial biosynthesis of alginates. *J chem Technol Biotechnol* **85**: 752-759.

Heinrich, P. C., H. Steffen, P. Janser & O. Wiss, (1972) Studies on the reconstitution of apotransketolase with thiamine pyrophosphate and analogs of the coenzyme. *Eur J Biochem* **30**: 533-541.

Hitchins, V. M. & H. L. Sadoff, (1973) Sequential metabolic events during encystment of *Azobacter vinelandii*. *J Bacteriol* **113**: 1273-1279.

Horecker, B. L., P. Z. Smyrniotis & H. Klenow, (1953) The formation of sedoheptulose phosphate. *J Biol Chem* **205**: 661-682.

Ignacimuthu, S., (2005) Basic bioinformatics. Harrow, Alpha Science International.

Iida, A., S. Teshiba & K. Mizobuchi, (1993) Identification and characterization of the *tktB* gene encoding a second transketolase in *Escherichia coli K-12*. *J Bacteriol* **175**: 5375-5383.

Jacobson, M. R., K. E. Brigle, L. T. Bennett, R. A. Setterquist, M. S. Wilson, V. L. Cash, J. Beynon, W. E. Newton & D. R. Dean, (1989) Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. *J Bacteriol* **171**: 1017-1027.

Janowicz, Z. A., M. R. Eckart, C. Drewke, R. O. Roggenkamp, C. P. Hollenberg, J. Maat, A. M. Ledeboer, C. Visser & C. T. Verrips, (1985) Cloning and characterization of the *DAS* gene encoding the major methanol assimilatory enzyme from the methylotrophic yeast *Hansenula polymorpha*. *Nucleic Acids Res* **13**: 3043-3062.

Josephson, B. L. & D. G. Fraenkel, (1969) Transketolase mutants of *Escherichia coli*. *J Bacteriol* **100**: 1289-1295.

Josephson, B. L. & D. G. Fraenkel, (1974) Sugar metabolism in transketolase mutants of *Escherichia coli*. *J Bacteriol* **118**: 1082-1089.

Kamada, N., A. Yasuhara, Y. Takano, T. Nakano & M. Ikeda, (2001) Effect of transketolase modifications on carbon flow to the purine-nucleotide pathway in *Corynebacterium ammoniagenes*. *Appl Microbiol Biotechnol* **56**: 710-717.

Kennedy, C., P. Rudnick, L. Melanie, MacDonald & M. THOYD, (2005) "Genus III. *Azotobacter Bejerinck* 1901, 567[AL] in: Bergey's mannual of Systematic Bacteriology, The Proteobacteria. New Yourk, Spinger.

Klug, W. S., S. M. Ward, C. Spencer & M. R. Cummings, (2006) Concepts of genetics. Pearson/Prentice Hall, Upper Saddle River, N.J.

Kochetov, G. A., (2001) Functional flexibility of the transketolase molecule. *Biochem (Mosc)* **66**: 1077-1085.

Lale, R., (2009) The *xylS/Pm* expression cassette: New functional insights and application potentials. Department of Biotechnology. Trondheim, NTNU.

Lin, L. P. & H. L. Sadoff, (1969) Chemical composition of A*zotobacter vinelandii* cysts. *J Bacteriol* **100**: 480-486.

Lin, T. Y. & W. Z. Hassid, (1966) Pathway of algnic acid synthesis in the marine brown alga, *Fucus gardneri Silva*. *J Biol Chem* **241**: 5284-5297.

Lindqvist, Y., G. Schneider, U. Ermler & M. Sundstrom, (1992) Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 A resolution. *EMBO J* **11**: 2373-2379.

Linker, A. & R. S. Jones, (1964) Polysaccharide resembling alginic acid from *Pseudomonas* micro-organism. *Nature* **204**: 187-188.

Littlechild, J., N. Turner, G. Hobbs, M. Lilly, A. Rawas & H. Watson, (1995) Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta Crystallogr D Biol Crystallogr* **51**: 1074-1076.

Lodish, H., (2008) Molecular cell biology. New York, W.H. Freeman and Company.

Loperfido, B. & H. L. Sadoff, (1973) Germination of *Azotobacter vinelandii* cysts: sequence of macromolecular synthesis and nitrogen fixation. *J Bacteriol* **113**: 841-846.

Marques, S., M. Manzanera, M. M. Gonzalez-Perez, M. T. Gallegos & J. L. Ramos, (1999) The *XylS*-dependent *Pm* promoter is transcribed *in vivo* by RNA polymerase with sigma 32 or sigma 38 depending on the growth phase. *Mol Microbiol* **31**: 1105-1113.

Mathew, M., (2008) Mechanistic studies of *Escherichia coli* transketolase. Department of Chemistry. UK, University of Exeter.

McPherson, M. J. & S. G. Møller, (2006) PCR. New York, Taylor & Francis.

Muhammadi & N. Ahmed, (2007) Genetics of bacterial alginate: alginate genes distribution, organization and biosynthesis in bacteria. *Curr Genomics* **8**: 191-202.

NEB, (2008) Bovine Serum Albumin (BSA). Source: New England BioLab. available at http://www.neb.com/nebecomm/products/productB9001.asp Downloaded 26.09.2008.

NEB, (2009) RbCl transformation protocol. Source: New England BioLab. available at http://www.neb.com/nebecomm/tech_reference/gene_expression/RbCl_protocol.asp Downloaded 15.01.2010.

NEB, (2010a) 1 kb DNA ladder. Source: New England BioLab. available at http://www.neb.com/nebecomm/products/productN3232.asp Downloaded 20.09.2010.

NEB, (2010b) Prestained Protein Ladder, Broad Range (10-230 kDa).Source: New England BioLab. available at http://www.neb.com/nebecomm/products/productP7710.asp Downloaded 15.10.2010.

Nicholl, D. S. T., (2002) An introduction to genetic engineering. Cambridge, Cambridge University Press.

Parente E., M. C., M Aquino and F Clementi, (1998) Alginate production by *Azotobacter vinelandii* DSM576 in batch fermentation *J. microbiol. and biotechnol.* **20**: 171-176.

Pelicic, V., J. M. Reyrat & B. Gicquel, (1996) Expression of the *Bacillus subtilis sacB* gene confers sucrose sensitivity on mycobacteria. *J Bacteriol* **178**: 1197-1199.

Péna, C., M. A. Trujillo-Roldan & E. Galindo, (2000) Influence of dissolved oxygen tension and agitation speed on alginate production and its molecular weight in cultures of *Azotobacter vinelandii*. *Enz Microb Technol* **27**: 390-398.

Pindar, D. F. & C. Bucke, (1975) The biosynthesis of alginic acid by *Azotobacter vinelandii*. *Biochem J* **152**: 617-622.

Promega, (2007) Technical Bullentin: Wizard Plus SV Minipreps DNA Purification System. Midison, Promega.

Pulakat, L., S. H. Lee & N. Gavini, (2002) Genome of *Azotobacter vinelandii*: counting of chromosomes by utilizing copy number of a selectable genetic marker. *Genetica* **115**: 147-158.

QIAGEN, (2008) QIAquick Spin Handbook. Qiagen.

Qurrat ul, A., S. Sharma, G. K. Khuller & S. K. Garg, (2003) Alginate-based oral drug delivery system for tuberculosis: pharmacokinetics and therapeutic effects. *J Antimicrob Chemother* **51**: 931-938.

Racker, E., de la Haba & I. G. leder, (1953) Thiamine pyrophosphate a coenzyme of transketolase. *J Amer Chem Soc* **74**: 1010.

Read, R. R. & J. W. Costerton, (1987) Purification and characterization of adhesive exopolysaccharides from *Pseudomonas putida* and *Pseudomonas fluorescens*. *Can J Microbiol* **33**: 1080-1090.

Reece, R. J., (2004) Analysis of Genes and Genomes. West Sussex, John Wiley & Sons, Ltd.

Rehm, B. H., (2010) Bacterial polymers: biosynthesis, modifications and applications. *Nat Rev Microbiol* **8**: 578-592.

Rehm, B. H., H. Ertesvåg & S. Valla, (1996) A new *Azotobacter vinelandii* mannuronan C-5-epimerase gene (*algG*) is part of an alg gene cluster physically organized in a manner similar to that in *Pseudomonas aeruginosa*. *J Bacteriol* **178**: 5884-5889.

Rehm, B. H. & S. Valla, (1997) Bacterial alginates: biosynthesis and applications. *Appl Microbiol Biotechnol* **48**: 281-288.

Remminghorst, U. & B. H. Rehm, (2006) Bacterial alginates: from biosynthesis to applications. *Biotechnol Lett* **28**: 1701-1712.

Robson, J. L. R. a. R. L., (1985) Isolation and properties of mutant of *Azotobacter chroococuum* defective in aerobic nitrogen fixation. *J general microbiol.* **131**: 1449-1458.

Russell, N. J. & P. Gacesa, (1988) Chemistry and biology of the alginate of mucoid strains of *Pseudomonas aeruginosa* in cystic fibrosis. *Mol Aspects Med* **10**: 1-91.

Sabra, A.P. Zeng, S.Sabry, S.Omar & W. D. Deckwer, (1999) Effect of phosphate and oxygen concentrations on alginate production and stoichiometry of metabolism of *Azotobacter vinelandii* under microaerobic conditions. *Appl Microbiol Biotechnol* **52**: 773-780.

Sabra, W. & A. P. Zeng, (2009) Microbial Production of alginates: Physiology and process aspects. In: B.H.A Rehm (ed) Microbiology Monography 13, Spinger, Berlin.

Sabra, W., A. P. Zeng & W. D. Deckwer, (2001) Bacterial alginate: physiology, product quality and process aspects. *Appl Microbiol Biotechnol* **56**: 315-325.

Sabra, W., A. P. Zeng, H. Lunsdorf & W. D. Deckwer, (2000) Effect of oxygen on formation and structure of *Azotobacter vinelandii* alginate and its role in protecting nitrogenase. *Appl Environ Microbiol* **66**: 4037-4044.

Sadoff, H. L., (1975) Encystment and germination in *Azotobacter vinelandii. Bacteriol Rev* **39**: 516-539.

Schneider, G. & Y. Lindqvist, (1998) Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis. *Biochim Biophys Acta* **1385**: 387-398.

Schaaff-Gerstenschlager, I. & F. K. Zimmermann, (1993) Pentose-phosphate pathway in *Saccharomyces cerevisiae*: analysis of deletion mutants for transketolase, transaldolase, and glucose 6-phosphate dehydrogenase. *Curr Genetics* **24**: 373-376.

Sci-Ed, s., (2007) Clone Manager Basic. Source: The Science and Education software.available at http://www.scied.com/pr_cmpro.htm Downloaded 3.10.2010.

Setubal, J. C., P. dos Santos, B. S. Goldman, H. Ertesvag, G. Espin, L. M. Rubio, S. Valla, N. F. Almeida, D. Balasubramanian, L. Cromes, L. Curatti, Z. Du, E. Godsy, B. Goodner, K. Hellner-Burris, J. A. Hernandez, K. Houmiel, J. Imperial, C. Kennedy, T. J. Larson, P. Latreille, L. S. Ligon, J. Lu, M. Maerk, N. M. Miller, S. Norton, I. P. O'Carroll, I. Paulsen, E. C. Raulfs, R. Roemer, J. Rosser, D. Segura, S. Slater, S. L. Stricklin, D. J. Studholme, J. Sun, C. J. Viana, E. Wallin, B. Wang, C. Wheeler, H. Zhu, D. R. Dean, R. Dixon & D. Wood, (2009) Genome sequence of *Azotobacter vinelandii,* an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J Bacteriol* **191**: 4534-4545.

Sigma-Aldrich, (1994) Enzymatic assay of Transketolase. Sigma-Adrich.

Sletta, H., A. Nedal, T. E. Aune, H. Hellebust, S. Hakvag, R. Aune, T. E. Ellingsen, S. Valla & T. Brautaset, (2004) Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. *Appl Environ Microbiol* **70**: 7033-7039.

Smidsrød, O. & K. I. Draget, (1996) Chemistry and physical properties of alginates. *Carbohydr. Europe* **14**: 6-13.

Socolofsky, M. D. & O. Wyss, (1961) Cysts of *Azotobacter*. *J Bacteriol* **81**: 946-954.

Socolofsky, M. D. & O. Wyss, (1962) Resistance of the *Azotobacter* cyst. *J Bacteriol* **84**: 119-124.

Sprenger, G. A., U. Schorken, G. Sprenger & H. Sahm, (1995) Transketolase A of *Escherichia coli K12*. Purification and properties of the enzyme from recombinant strains. *Eur J Biochem* **230**: 525-532.

Steigedal, M., (2006) The *Azotobacter vinelandii* mannuronan C5-epimerases: their biological functions and new tools useful for their future *in vivo* biotechnological application. Department of Biotechnology,Trondheim, NTNU.

Steigedal, M., H. Sletta, S. Moreno, M. Maerk, B. E. Christensen, T. Bjerkan, T. E. Ellingsen, G. Espin, H. Ertesvåg & S. Valla, (2008) The *Azotobacter vinelandii* AlgE mannuronan C-5-epimerase family is essential for the in vivo control of alginate monomer composition and for functional cyst formation. *Environ Microbiol* **10**: 1760-1770.

Stevenson, L. H. & M. D. Socolofsky, (1966) Cyst formation and poly-beta-hydroxybutyric acid accumulation in *Azotobacter*. *J Bacteriol* **91**: 304-310.

Stokke, B. T., O. Smidsrod, P. Bruheim & G. Skjåk-Bræk, (1991) Distribution of uronate residues in alginate chains in relation to alginate gelling properties. *Macromol* **24**: 4637-4645.

Thornton, J., C. Orengo & D. Jones, (2003) Bioinformatics : genes, proteins and computers. Oxford, BIOS Scientific Publishers Ltd.

Valla, H. E. a. S., (1998) Biosynthesis and applications of alginates. *Pol Degrad Stab* **59**: 85-91.

Valla, S., J. Li, H. Ertesvåg, T. Barbeyron & U. Lindahl, (2001) Hexuronyl C5-epimerases in alginate and glycosaminoglycan biosynthesis. *Biochimie* **83**: 819-830.

Vectronbiosolutions, (2010) The *Pm/xylS* expression systmen. Source: Vectron biosolutions. available at http://vectronbiosolutions.com/info.php?id=13 Downloaded 20.10.2010.

Wamelink, M. M., E. A. Struys & C. Jakobs, (2008) The biochemistry, metabolism and inherited defects of the pentose phosphate pathway: a review. *J Inherit Metab Dis* **31**: 703-717.

Watson, Baker, Bell, Gann, Levine & Losick, (2004) Molecular biology of the genes. USA, Pearson Edication Inc.

Wikner, C., U. Nilsson, L. Meshalkina, C. Udekwu, Y. Lindqvist & G. Schneider, (1997) Identification of catalytically important residues in yeast transketolase. *Biochem* **36**: 15643-15649.

Wilson, K. J., A. Sessitsch, J. C. Corbo, K. E. Giller, A. D. Akkermans & R. A. Jefferson, (1995) beta-Glucuronidase (GUS) transposons for ecological and genetic studies of rhizobia and other gram-negative bacteria. *Microbiol* **141 (Pt 7)**: 1691-1705.

Winther-Larsen, H. C., K. D. Josefsen, T. Brautaset & S. Valla, (2000) Parameters affecting gene expression from the *Pm* promoter in gram-negative bacteria. *Metab Eng* **2**: 79-91.

Wyss, O., M. G. Neumnn & M. D. Socolofsky, (1961) Development and germination of the *Azotobacter* cyst. *J Biophys Biochem Cytol* **10**: 555-565.

Zhao, J. & C. J. Zhong, (2009) A review on research progress of transketolase. *Neurosci Bull* **25**: 94-99.

Zvelebil, M. J. & J. O. Baum, (2008) Understanding bioinformatics. New York, Garland Science.

Appendixes

Appendix A List of Abbreviations
Appendix B Plasmid maps
Appendix C DNA standard for gel electrophoresis
Appendix D The SDS-PAGE standard marker
Appendix E Recognition sites for endonucleases enzymes
Appendix F Data and standard curve for protein concentration
Appendix G Gene sequences
Appendix H Alignment for PCR products
Appendix I Graphs of activity measuring for the transketolase proteins.

Appendix A List of abbreviations

aa	amino acids
Amp	ampcilline
BA	Burks agar
BLASTP	protein BLAST
BLASTN	nucleotides BLAST
BLASTX	Search protein database using a translated nucleotide query
BM	Burks medium
BSA	Bovin serum albumin
Ca^{2+}	Calcium
DNA	deoxyribonucleic acid
ddNTP	dideoxynucleoside triphosphates
dNTP	deoxyribonucleotides
EDTA	ethylenediaminetetraacetic acid
GDP	guanosine diphosphate
H^{+}	Hydrogen
kb	kilo base
KEGG	Kyoto Encyclopedia of Genes and Genomes)
LA	Luria-Bertani agar
LB	Luria-Bertani medium
MOPS	The 3-(N-morpholino)-propane-sulfonic acid
NADH	Nicotinamide adenine dinucleotide
nt	nucleotide
OD	optical density
bp	Base pair
PCR	Polymerase chain reaction
pH	The Power of Hydrogen
PHB	poly- β-hydroxybutyric acid
SOC	Super Optimal Broth with Catabolite repression
Sp^{R}	spectinomycin resistance
TCA	Trichloroacetic acid
ThDP	thiamine diphosphate
UV	Ultraviolet

Appendix B Plasmid maps

Figure B1 a-w, shows the parental plasmids and the designated plasmids obtained in this work. pHE216, pHE263, pHE272, pAKV1, pMK2, pMK3, pMK4, pMK5, pMK6, pMK7, pMK8, pMK9, pMK10, pMK11, pMK12, pMK13, pMK14, pMK15, pMK16, pMK17, pHE303, pHE304 and pME2.
The recognition sites for restriction enzymes used are shown on the plasmid map, as well as genes characterizing each plasmid.

a)

b)

c)

d)
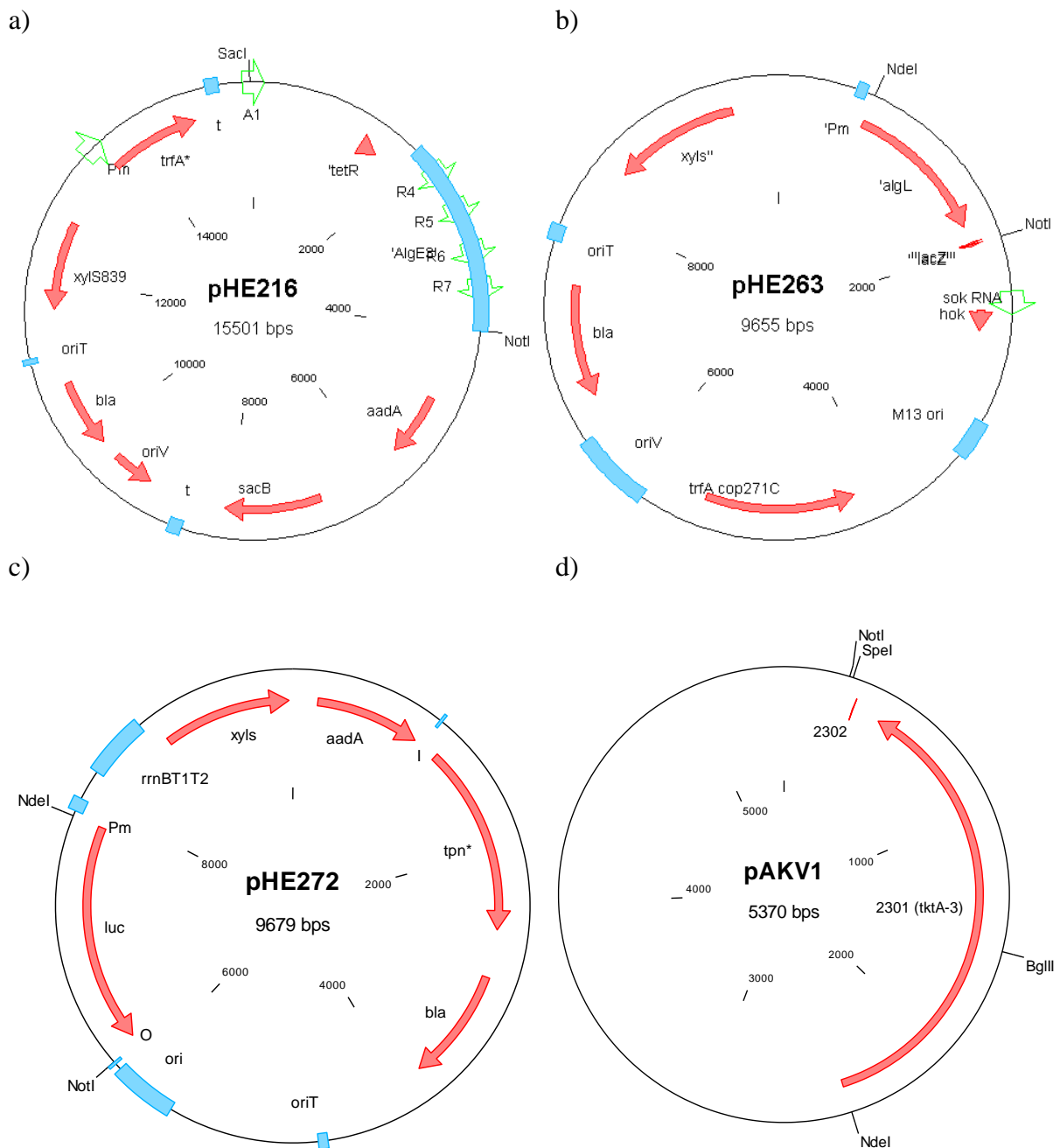
Figure B1 plasmid physical map with their restriction sites a) pHE216 b) pHE263 c)pHE272 and d) pAKV1 .

e) pMK2
NdeI
Pm
rrnBT1T2
Xyls'
pMK2
6066 bps
6000
5000
1000
bla
4000
2000
3000
M13ori
ori
O
NotI

f) pMK3
xyls
aadA
rrnBT1T2
I
NdeI
Pm
tpn*
tktA-1
pMK3
10131 bps
10000
8000
2000
4000
bla
6000
O
ori
oriT
NotI

g) pMK4
PciI
O
'tktA-1
SacI
tktA-2''
pMK4
5236 bps
5000
1000
4000
2000
3000
'''lacZ''

h) pMK5
''''lacZ''''
tktA-2''
aadA
'tktA-1
'''''lacZ'''''
t
trfA*
pMK5
13499 bps
12000
2000
sacB
10000
4000
Pm
t
EcoRV
8000
6000
xylS839
oriV
bla
oriT
oriT
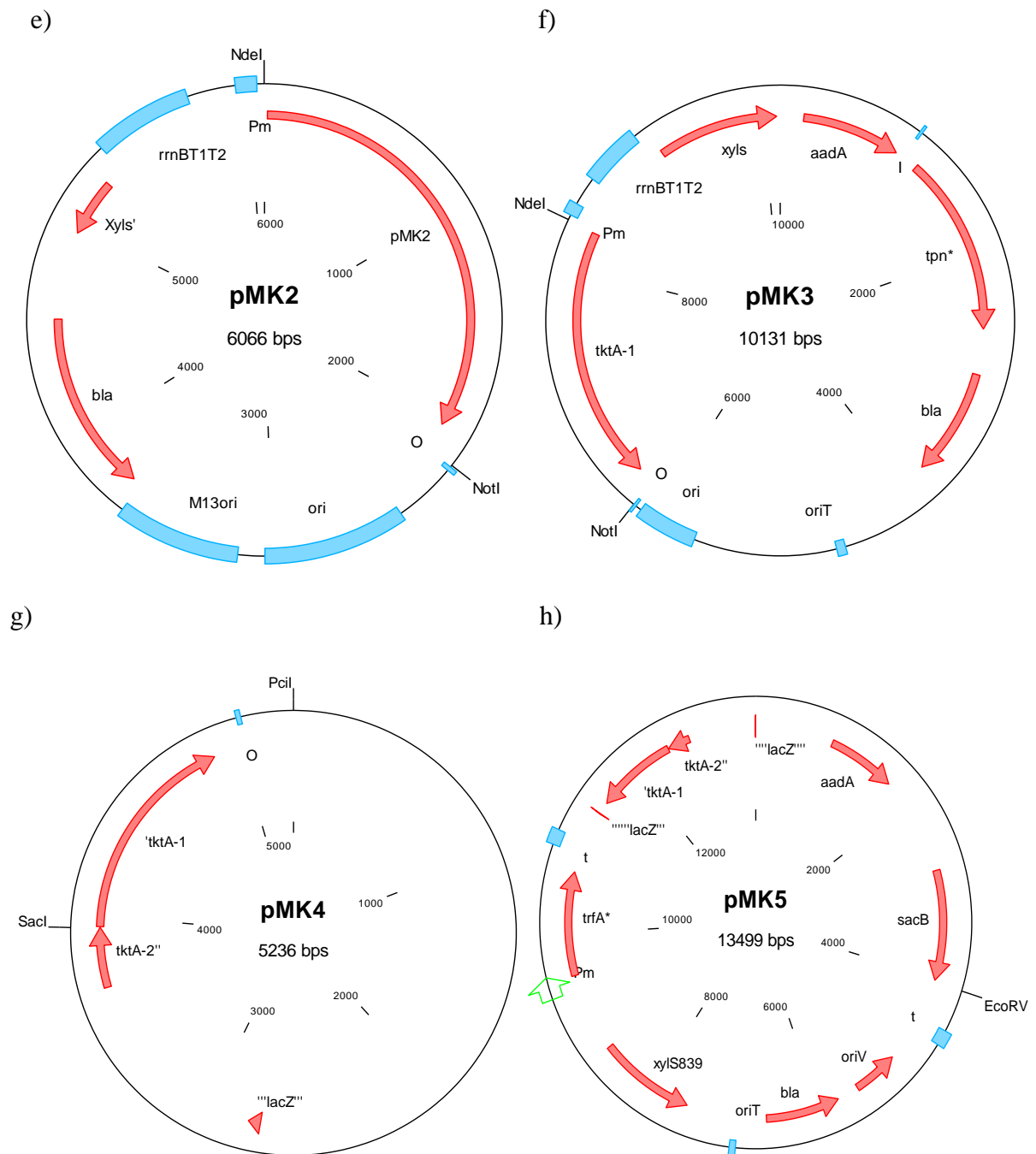
Figure B1 plasmid physical map with their restriction sites, e) pMK2, f) pMK3,g) pMK4 and h) pMK5.

92
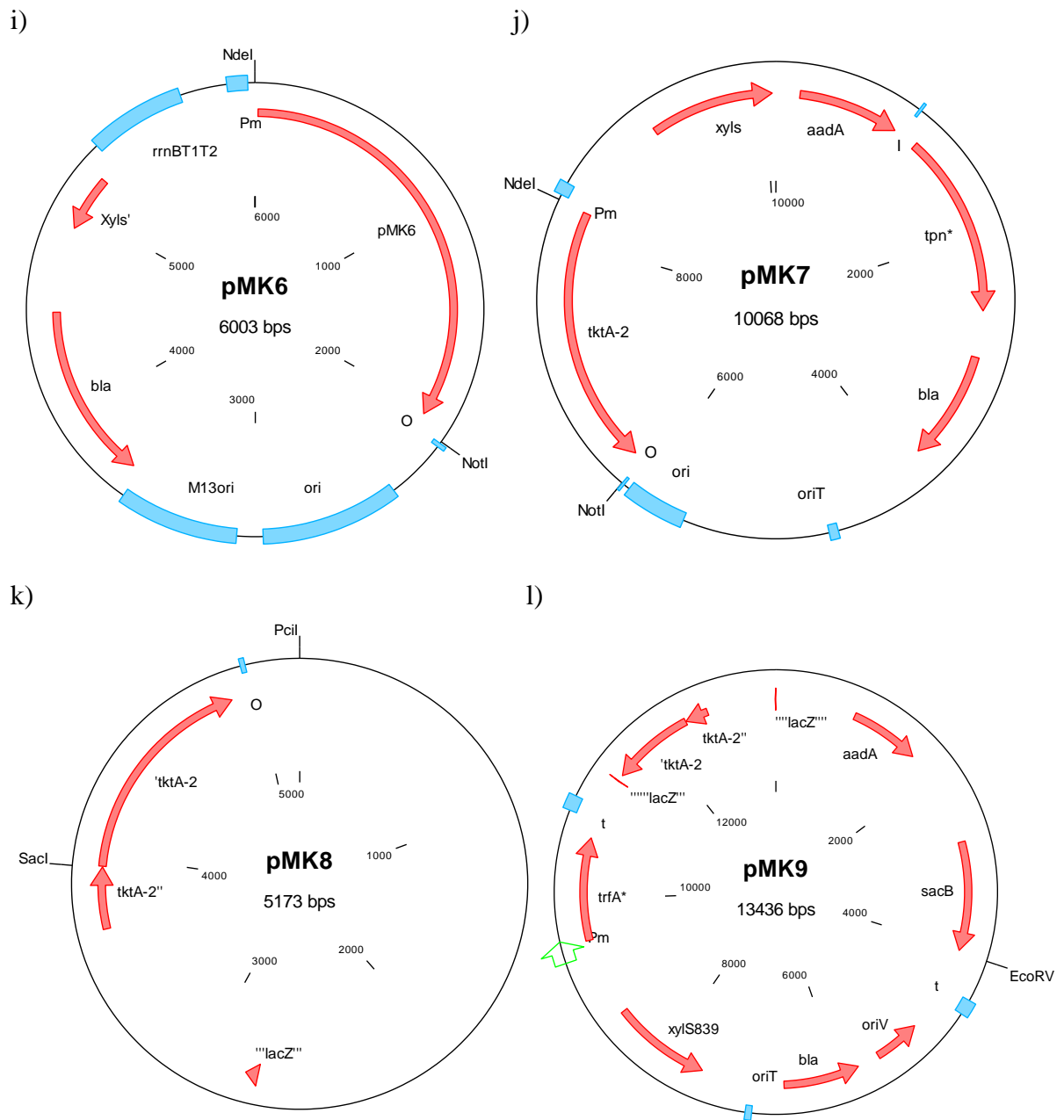
Figure B1 plasmid physical map with their restriction sites i) pMK6, j) pMK7, k) pMK8 and l) pMK9.
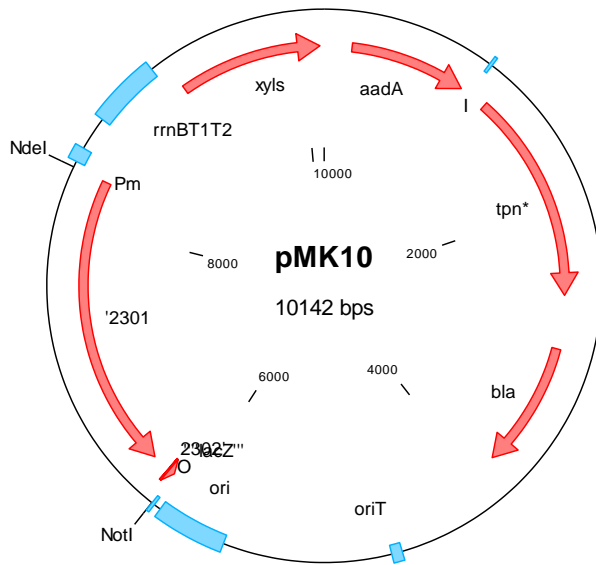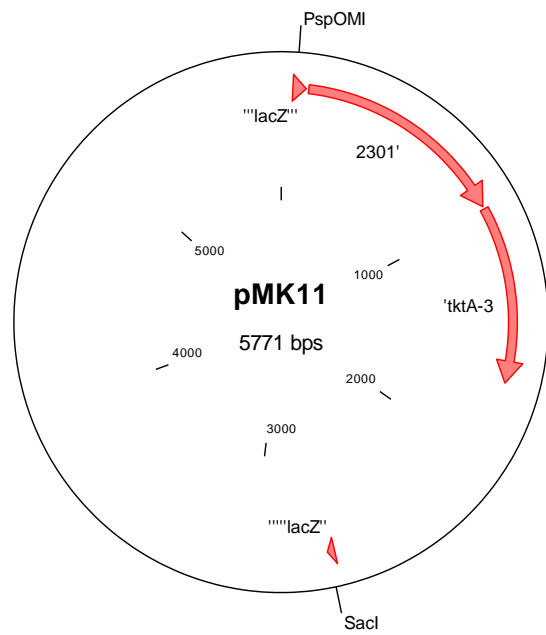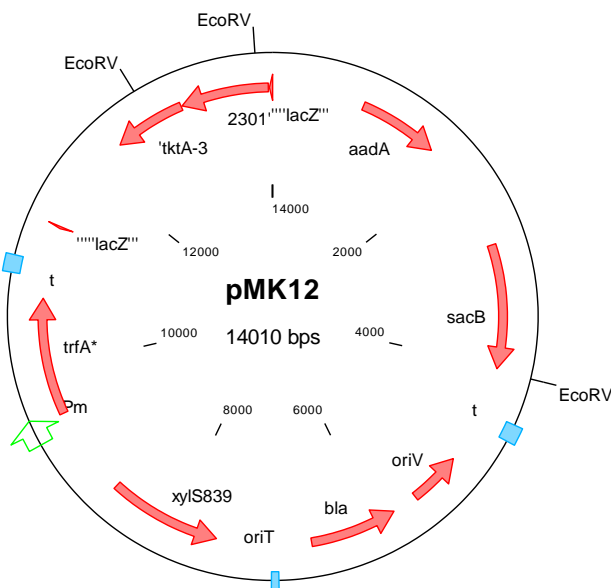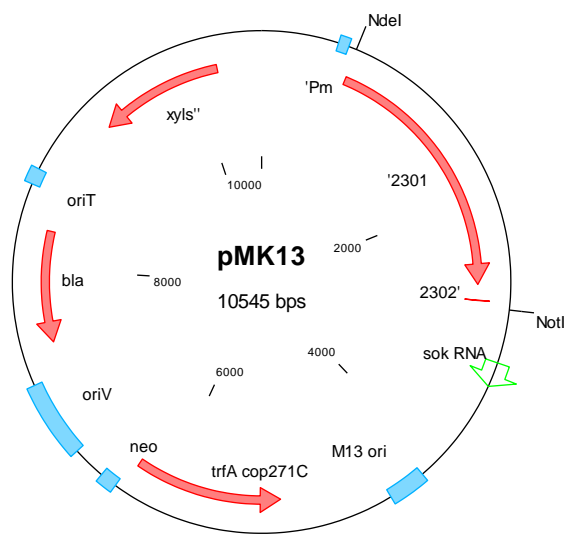
Figure B1 plasmid physical map with their restriction sites m) pMK10, n) pMK11, o) pMK12 and p) pMK13.
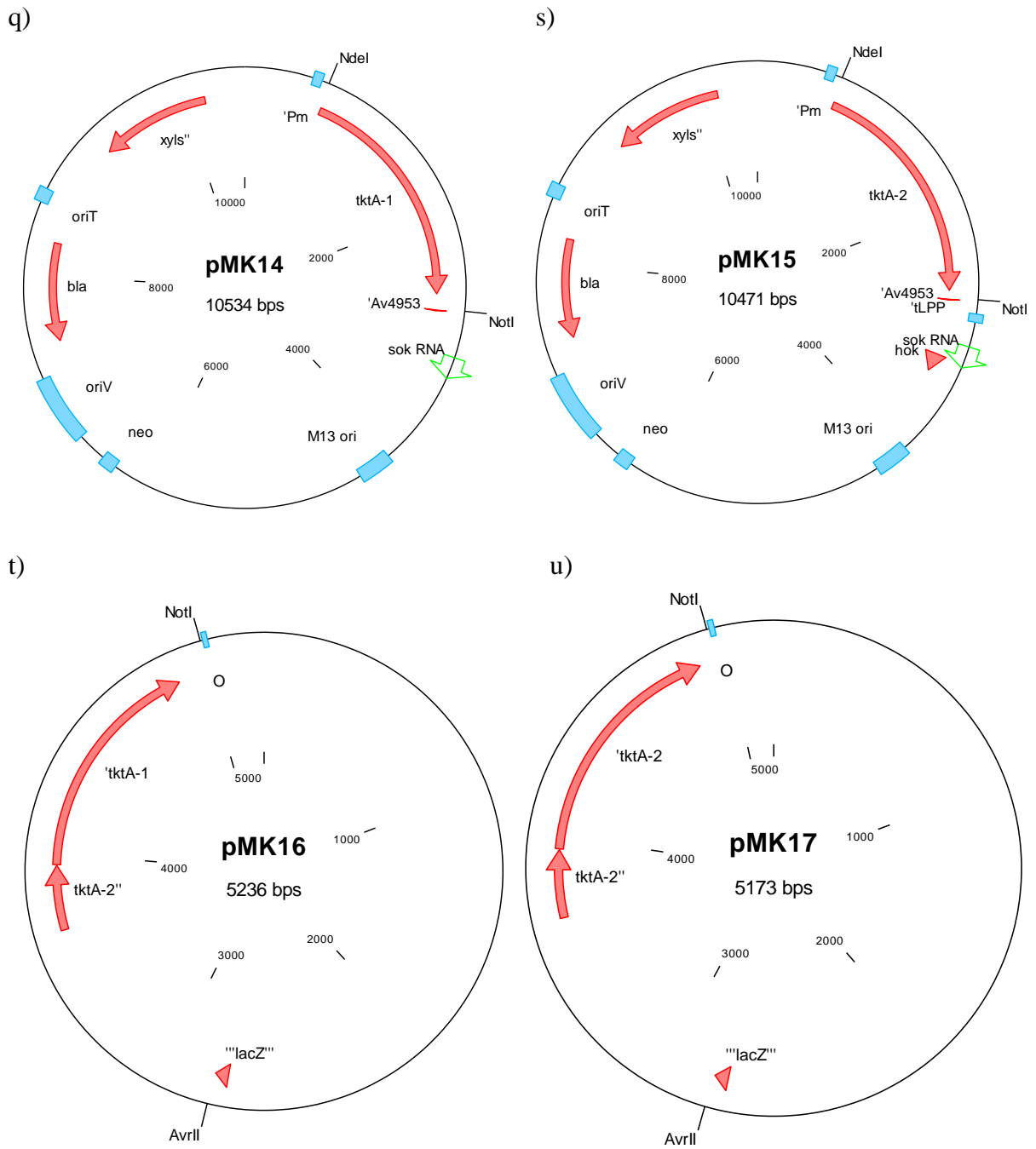
Figure B1 plasmid physical map with their restriction sites q) pMK14, s) pMK15, t) pMK16 and u) pMK17.

v)
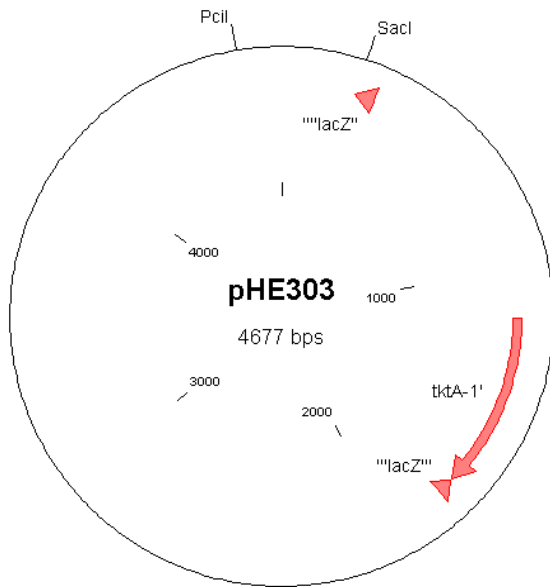
w)
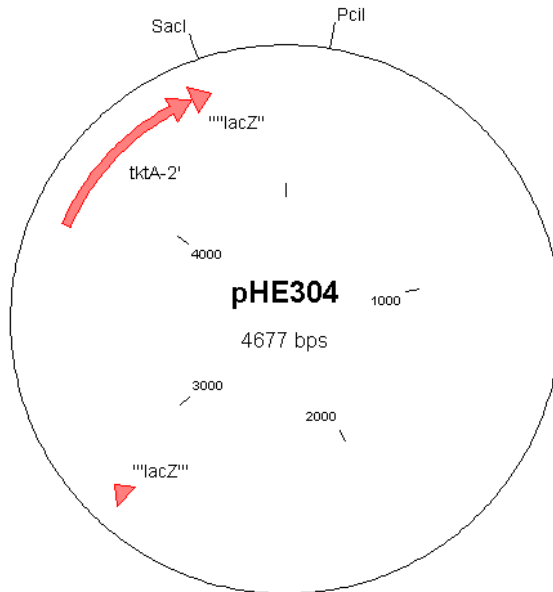
x)
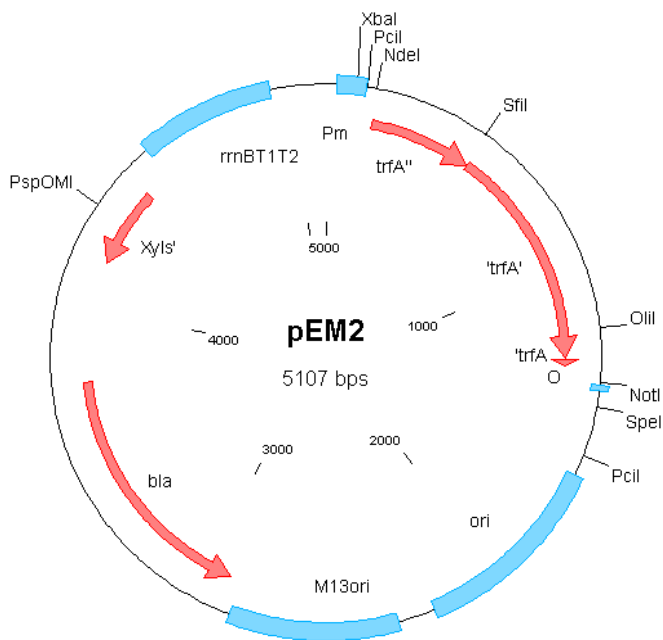
Figure B1 plasmid physical map with their restriction sites v) pHE303 and w) pHE304 x) pEM2.

Appendix C DNA standard for gel electrophoresis

The DNA standard used in the determination of DNA size fragments after separation with gel electrophoresis were TriDye™ 1 kb DNA ladder (Figure C1) and N3200L 2-logg ladder (Figure C2) from New England Biolab (NEB). The concentration of DNA fragments were estimated using these DNA standards.
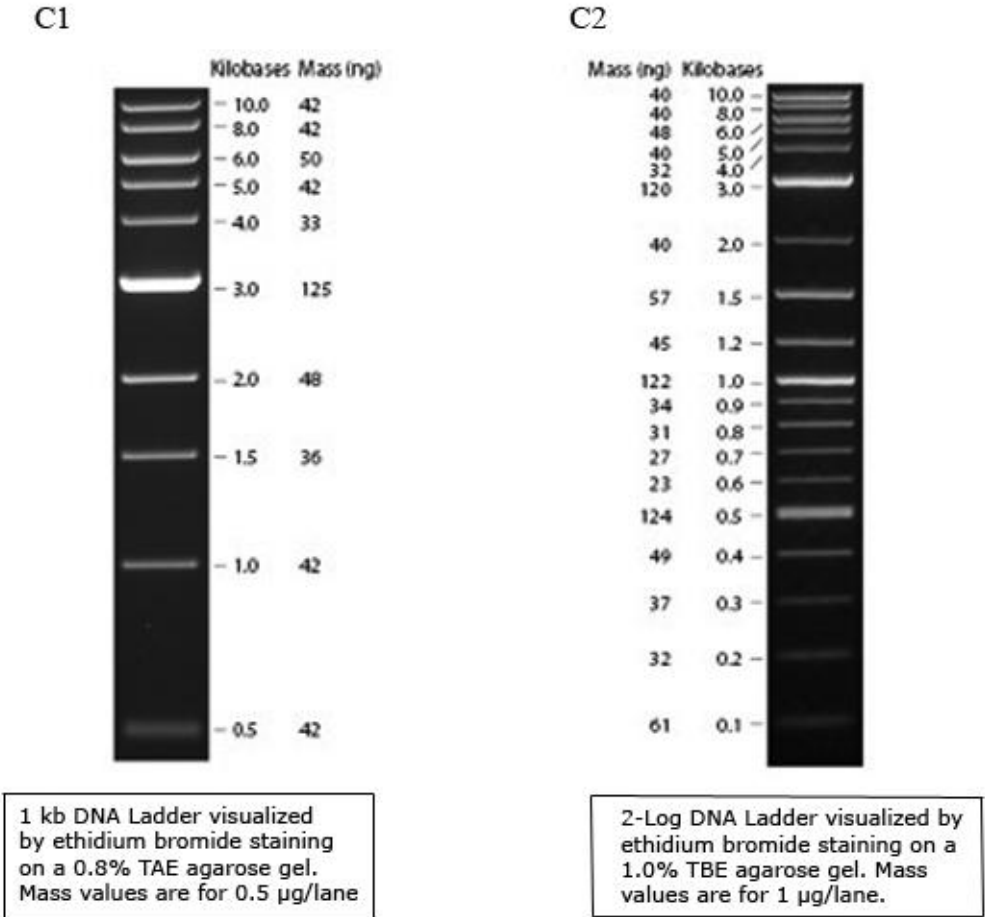
Figure C 1) TriDye™1 kb DNA ladder from NEB C2) N3200L 2-logg ladder from NEB (NEB, 2010a).

Appendix D The SDS-PAGE standard marker

The SDS-PAGE standard market is used in the determination of protein size fragments after separation with polyacrylamid gel electrophoresis was P7710S 230 kDa protein ladder (Figure D) from New England Biolab (NEB).
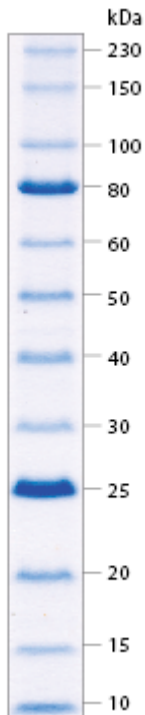
Invitrogen Novex 10-20% Tris-glycine SDS-PAGE gel. **Concentration:** 0.2 mg/ml

Figure D Prestained P7710S Protein Ladder, Broad Range (10-230 kDa) from NEB (NEB, 2010b).

Appendix E Recognition sites for endonucleases enzymes

Table E. shows the restriction enzymes used in this work with their recognition sites.

Table E The restriction enzymes (RE) with their recognition site, The down and up-pointing triangles indicate the site of digestion.

Restriction enzymes	Recognition site
NotI	5´...GC▼GGCCGC...3´ 3´...CGCCGG▲CG...5´
NdeI	5´...CA▼TATG...3´ 3´...GTAT▲AC...5´
SacI	5´...GAGCT▼C...3´ 3´...C▲TCGAG...5´
EcoRI	5´...G▼AATTC...3´ 3´...CTTAA▲G...5´
DraIII	5´...CACNNN▼GTG...3´ 3´...GTG▲NNNCAC...5´
AvrII	5´...C▼CTAGG...3´ 3´...GGATC▲C...5´
SalI	5´...G▼TCGAC...3´ 3´...CAGCT▲G...5´
SpeI	5´...A▼CTAGT...3´ 3´...TGATC▲A...5´
BglII	5´...A▼GATCT...3´ 3´...TCTAG▲A...5´
PspOMI	5´...G▼GGCCC...3´ 3´...CCCGG▲G...5´
XhoI	5´...C▼TCGAG...3´ 3´...GAGCT▲C...5´
EcorV	5´...GAT▼ATC...3´ 3´...CTA▲TAG...5´
PciI	5´...A▼CATGT...3´ 3´...TGTAC▲A...5´

Appendix F Data and standard curve for protein concentration estimation

Protein concentration in the cell free extracts from *S 17.1* and *S17.1 pir* with addition of (0.5 mM final concentration) of m-toluic acid as inducer was measured using Bio-Rad Protein Assay. Table F.1 shows the data obtained from the measurement of absorbance at 595 nm of BSA (the standard sample). Table F.2 shows the measurements of the *tktA2* and *tktA3* protein absorbance at 595 nm.

Table F.1 The absorbance of standard (BSA) at OD$_{595}$.

Standard (BSA µg/ml)	I	II	III
0	0.596	0.609	0.607
0.8	0.639	0.628	0.610
1.6	0.671	0.690	0.720
3.2	0.740	0.746	0.762
4.8	0.814	0.860	0.820
6.4	0.860	0.899	0.900

Table F.2 The absorbance of TktA2 and TktA3 in cell free extract at OD$_{595}$.

Proteins	I	II	III
TktA2 (3 µl/ml)	0.689	0.690	0.693
TktA3 (5 µl/ml)	0.899	0.901	0.878

Table F.2 The absorbance of TktA2 and TktA3 in fraction obtained from FPLC at OD$_{595}$.

Proteins	I	II	III
TktA2 (3 µl/ml)	0.879	0.882	0.888
TktA3 (5 µl/ml)	0.766	0.765	0.761

The standard curve used to estimate the protein concentration in the cell free extract of the TktA2 and TktA3 proteins. The standard curve plot base on data in table F.1.
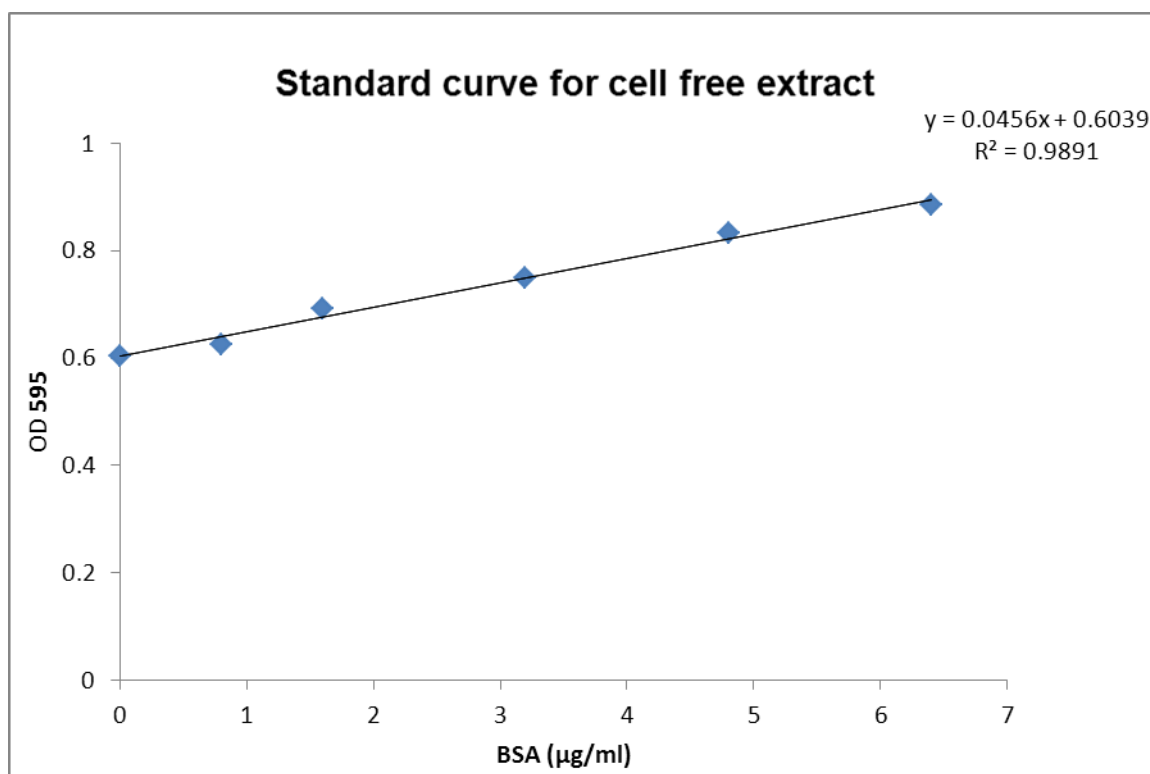
Figure F.1 Standard curve for the estimation of protein concentration in the cell free extract and fraction obtained from FPLC.

5 µl of cell free extract of TktA3 and 3 µl of cell cell free extract of TktA2 in total volume of 1000 µl in cuvettes gave the absorbances described in table F.2. that were found in the area of the standard curve from 0.6-0.9 nm. From the average of absorbance measurements, and the equation of linearity obtained from the standard curve the concentration of each sample is calculated.

Table F.3 shows the absorbances of 5 µl and 3 µl of fractions obtained from the FPLC in total volume of 1000 µl in cuvettes. The concentration of each sample is calculated from the equation of linearity for the standard curve.

Calculation of protein amount in TktA2 cell free extract;

$Y = 0.690$
$X = (0.690-0.6039)/0.0456$
$X = 1.8$ µg/ml

Calculation of protein amount in TktA3 cell free extract;

$Y = 0.892$
$X = (0.892-0.6039)/0.0456$
$X = 6.3$ µg/ml

Calculation of protein amount in TktA2 x1 fractions of FPLC;

Y= 0.883
X= (0.883-0.6039)/0.0456
X= 5.0 µg/ml

Calculation of protein amount in TktA3 x1 fraction of FPLC;

Y= 0.764
X= (0.764-0.6039)/0.0456
X= 3.5 µg/ml

The protein concentration in sample = (µl total volume assay)/ (µl volume of sample used) x the protein concentration calculated from the standard curve.

Protein concentration in TktA2 cell free extract;

(1000 µl / 5 µl) x 1.8 µg/ml = 0.360 mg/ml
0.460 mg/ml x 5 ml = 1.8 mg

Protein concentration inTktA3 cell free extract;

(1000 µl / 5 µl) x 6.3 µg/ml = 1.260 mg/ml
1.260 mg/ml x 5 ml = 6.3 mg

Protein concentration in TktA2 FPLC x1 fractions;

(1000 µl / 5 µl) x 5.0 µg/ml = 1.0 mg/ml
1.0 mg/ml x 5 ml = 5.0 mg

Protein concentration in TktA3 FPLC x1 fractions;

(1000 µl / 5 µl) x 3.5 µg/ml = 0.700 mg/ml
 0.700 mg/ml x 5 ml = 3.5 mg

Appendix G Gene sequences and the secondary structure prediction

G.1 The gene sequences of *tktA1*, *tktA2* and *tktA3* blasted in NCBI and their similarity to *P. aeruginosa* shown in percentage as well as the length of each sequence.
G.2 The multiple sequence alginment of the three transketolase gene shows the similarity and difference in these three genes.
G.3 The multiple alginment of the three transketolase gene to both *tktA* and *tktB* from *E. coli*, the organism of most studied transketolase in prokaryote.
G.4 The multiple alignment for each gene and deletion mutant constructed for this work

G.1.1 BLAST result of *tktA1* gene on NCBI shows the best hits against *Pseudomonas aeruginosa* with 78% similarity.

```
 Score = 1024 bits (554), Expect = 0.0
 Identities = 1562/2021 (78%), Gaps = 179/2021 (8%)
 Strand=Plus/Plus

Query  47      GCCAATGCCATCCGCGTC-CTGAGCATGGATGCGGTG--GAACGGGCGAACTCCGGCCAT  103
               |||||||||||||| | | ||||||||||||| ||| ||| || ||| ||||||
Sbjct  665373  GCCAATGCCATCCGTG-CACTGAGCATGGATGCCGTGCAGAA--AGCCAACAGCGGCCA-  665428

Query  104     CCC-GGCGCCCCCATGGGCATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGC-  161
               ||| |||||||| ||||||||||||||||||||||| || |||||| || || ||| |||
Sbjct  665429  CCCGGGCGCCCCGATGGGCATGGCCGATATCGCCGAGGTCCTCTGGCGCGACTACATGCA  665488

Query  162     GCCACAACCCGGGCAACCC-CCGCTGGCCCGACCGCGACCGCTTCGTGCTCTCCAACGGG  220
               | |||||||| ||||||| | | ||| || |||||||||||||||||||| ||||||||
Sbjct  665489  G-CACAACCCGAGCAACCCGCAG-TGGCCCAACCGCGACCGCTTCGTGCTGTCCAACGGC  665546

Query  221     CACGGCTCGATGCTGCTCTACAGCCTGCTGCACCTCTCCGGCTACGA-CTTGCCGGTCGA  279
               ||||||||||||||| ||||| ||||||||||||||| |||| ||||| || | | ||||
Sbjct  665547  CACGGCTCGATGCTGATCTACTCCCTGCTGCACCTCACCGGGTACGACCTCGGC-ATCGA  665605

Query  280     GGAGCTCAAGCGCTTCCGCCAGTTGC-ATTCGCGCACTCCCGGG-CACCCGGAGTACGGC  337
               ||| || ||| |||||||||||||| | | | ||||||||| ||||| ||||||| ||||||
Sbjct  665606  GGACCTGAAGAACTTCCGCCAGCT-CAACTCGCGCAC-CCCGGGTCACCCGGAATACGGC  665663

Query  338     TACA-CGCCGGGGGTGGAGACCACCACCGGTCCGCTCGGCCAGGGGC-TGGCCAACGCCG  395
               |||| |||| |||| || |||||||||||||||||||||||||| |||| | ||||| || |
Sbjct  665664  TACACCGCC-GGCGTCGAGACCACCACCGGTCCGCTCGGCCA-GGGCATCGCCAATGCGG  665721

Query  396     TCGGCATGGCCCTGGCCGAG-CGGCTGCTGGCGGCGGACTTCAACCGC-CCGGGACAGGC  453
               | ||||||||| ||||| ||| || | |||||| || | |||||||||| | || || ||
Sbjct  665722  TGGGCATGGCGCTGGCGGAGAAGG-TCCTGGCCGCCCAGTTCAACCGCGAC-GGCCACGC  665779

Query  454     GATCGTCGACCACCGCACCTACGCCTTCGTCGGCGACGGCTGCCTGATGGAGGGCATCTC  513
               | | |||||||||| |||||||||||| |||||||||||| ||| ||||||| ||||| ||
Sbjct  665780  GGTGGTCGACCACTACACCTACGCCTTCCTCGGCGACGGTTGCATGATGGAAGGCATTTC  665839

Query  514     CCACGAGGTCTG-CTCGCTGGCCGGCACCCTGGGACTCG-GCAAGCTGATCGTGTTCTAC  571
               ||| ||||||| | ||||||||||||||||||| | || | |||||||||| ||||||
Sbjct  665840  CCATGAGGTC-GCCTCGCTGGCCGGCACCCTGCGCCT-GAACAAGCTGATCGCCTTCTAC  665897

Query  572     GACGACAACGGCATCTCCATCGACGGTGAGGTCG-CGGGCTGGTTCAGCGACGACACGCC  630
               |||||||||||||||||||||||||||| ||||| ||| ||||||||||| |||||| ||
Sbjct  665898  GACGACAACGGCATCTCCATCGACGGCGAGGTCCACGG-CTGGTTCACCGACGACACCCC  665956

Query  631     GGCGCGCTTCCGCG-CCTACGGCTGGCAGGTCG-TC-G-AGGCGGTCGACGGCCACGACG  686
               | |||||| | ||||||||||||||||||||| |||| | | |||||||| ||||||||
Sbjct  665957  GAAGCGCTTC-GAGGCCTACGGCTGGCAGGT-GATCCGCAAT-G-TCGACGGGCATGACG  666012

Query  687     CC-AGCGCGCTCGAGGCGGCGC-TGG---CGGCG-GCCAAG-GCCGATACGGAGCGGCCG  739
               || | || || || | || | | || | ||| || ||| || || | ||| | |
Sbjct  666013  CCGA-CGAGATCAAGACCGC-CATCGATACCGCGCGC-AAGAGC-GAC-C--AGC--C-G  666062

Query  740     ACGCTGATCTGCTGCAAGACCACCATCGGCTTCGGCGCGCCGACCAAGGCCGG-CA-GCC  797
               || | |||||||||||||||| || ||| || ||||| || |||| |||| || | |
Sbjct  666063  ACCCTGATCTGCTGCAAGACCGTGATCGGCTTCGGCTCGCCGAACAAG-CAGGGCAAGGA  666121
```

103

```
Query  798    ACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGATCGCCGCGGCCCGCGAGGCGCTGG  857
              | ||  ||||||||||| ||||||||| ||| |||||||||||| ||||||  ||||||||
Sbjct  666122  A-GAGTGCCACGGCGCGCCGCTGGGCGCCGACGAGATCGCCGCGACCCGCGCCGCGCTGG  666180

Query  858    ACTGGCCGCACGCGCCC-TTCGAGGTGCCCG-CGCAGATC---GCCAGGGCCTGGGACGC  912
              ||||  ||||| | || ||||||| | |||| |||||||||   ||| | |  ||||||||
Sbjct  666181  GCTGGGAGCACG-GTCCGTTCGAGAT-CCCGGCGCAGATCTACGCC-GAG--TGGGACGC  666235

Query  913    GC-GCGAG-CGCGGGGCGGCGCTGGA-GCGGGACTGGCGACAA-CGCTTCGACGACTACG  968
              || | || | ||| || || || ||| || |||| |||| || ||| | |||| || ||||
Sbjct  666236  -CAGGGAAAC-CGGCGCCGCCCAGGAAGCCG-AGTGGA-ACAAGCGTTTCGCCGCCTAC-  666290

Query  969    C-G-CGGGCCTACCCGGCGCTGGCCGCCGAACTGGAGCGGCGCCTGGCCGGCGAGCTGCC  1026
              | | | || ||  |||| |||||||||||||||| ||| ||||||  |||||||||||||
Sbjct  666291  CAGGCCGCCC-ATCCGGAACTGGCCGCCGAACTGCTGCGCCGCCTGAAGGGCGAGCTGCC  666349

Query  1027   GGCCGACTGGggcgcgc-ggcgcgac-gcgct--gatcgccg--g-ctgccagcaggcgg  1079
              ||||||||| | ||| | ||| || ||  | ||| | ||| | | |||| || | ||
Sbjct  666350  GGCCGACTTCGCCGAGAAGGC-CG-CTGC-CTACG-TCGCCGATGTC-GCCAACAAG-GG  666403

Query  1080   cgcg-cgagctggcgACCCGCAAGAGCTC-GC-AGCAGGTC-CTGGAAGCGC-TG-GGCC  1133
              || | | | | |||||| || | || | | | | ||| | ||| | |||| ||||
Sbjct  666404  CGAGACCATC-GCC-AGCCGCAAG-GC-CAGCCAGAACG-CGCTG-AA-CGCCTTCGGCC  666456

Query  1134   CGCTGCTGCCGGAGCTGCTCGGCGGCTCGGCGGACCTGGC-GCCCTCGAACCTGACCC-G  1191
              ||||||||||||||||||||| || || |||||||| | | || |||||||||||||| |
Sbjct  666457  CGCTGCTGCCGGAGCTGCTCGGCGGTTCCGCCGACCTGGCCGGC-TCCAACCTGACCCTG  666515

Query  1192   CTGGTCCGGCTCCC-GCTCGGTCGGCGGC-GAGG-CGCCCGGGGGCAACTACATCCACTA  1248
               |||   ||||   |  || |||| ||| | |||||| |||||||||||||| | ||
Sbjct  666516  -TGGAAGGGCTGTAAGGGCG-TCAGCG-CTGACGACGCC-GCCGGCAACTACGTGTTCTA  666571

Query  1249   CGGGGTGCGCGAGTTCGGCATGAGCGCGATGATGAACGG-GCTGGCGCTGCACGGCGGTT  1307
              ||| |||||||||||||||||||||||||||  |||||||| | | || ||||||||||||
Sbjct  666572  CGGCGTGCGCGAGTTCGGCATGAGCGCGATCATGAACGGCG-TCGCCCTGCACGGCGGTT  666630

Query  1308   TCATCCCCTACGGCG-GCACCTTCCTGATGTTCATGGAATACGCCCGCAACGCCGTGCGC  1366
              |||| |||||||||||| ||||||||||| || |||||| ||||| ||||||||||||||||
Sbjct  666631  TCATTCCCTACGGCGCG-ACCTTCCTGATCTTCATGGAGTACGCGCGCAACGCCGTGCGC  666689

Query  1367   ATGGCCGCGCTGATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCCTG  1426
              ||| |||| |||||||||||  ||| | | ||| | |||||| |||||  ||||||||
Sbjct  666690  ATGTCCGCCCTGATGAAGCAGCGCGTGCTCTACGTGTTCACCCATGACTCCATCGGCCTC  666749

Query  1427   GGCGAGGACGGCCCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCGCCAGACGCCG  1486
              |||||||||||||||||||||||||||| ||| || || |||||||||||||| ||| |||
Sbjct  666750  GGCGAGGACGGCCCGACCCACCAGCCGATCGAGCAACTGGCCAGCCTGCGCCTGACCCCG  666809

Query  1487   AACCTGGAAACCTGGCGGCCCTGC-GACGAGA-CGGAGACCGCGGTGGCCTGGT-GCGCG  1543
              ||||||| |||||||||| ||||||| ||| || || || | |||||||||||  ||  |
Sbjct  666810  AACCTGGACACCTGGCG-CCCGGCCGATGCGGTCG-AGTCGGCGGTGGCCTGGAAGCATG  666867

Query  1544   GCGCTCGAACG-GCGGGACGGCCCGTCGGCGCTGGTGCTG-Tcgcggcagaacctgccgg  1601
              | | |||| || || || |||| ||||| ||||| | || | | |||||||||||||||||
Sbjct  666868  CCA-TCGAGCGTGCCG-ACGGTCCGTCCGCGCTGAT-CTTCTCCCGCCAGAACCTGCCG-  666923

Query  1602   ccgc-ggcgcgcgacgccgggcagcgcg-cggccatcgagcgcggcgccTATGTGCTGCT  1659
              | | | ||||||||||| || ||| | | | | ||||| |||||||| ||| || |||
Sbjct  666924  CATCAGGCGCGCGACGTCGCCCAG-GTGACCGACATCGCGCGCGGCGGCTACGTACTGAA  666982

Query  1660   CGACTGC-CCGG-GCACGCCGGAGCTGATCCTgctggcgtccggctccgaggt-ggcgct  1716
              ||||||| |||| || | | ||||| |||||||| | || |||| || || ||| || ||
Sbjct  666983  GGACTGCGCCGGCG-A-GCCGGAACTGATCCTGATCGCCACCGGTTCGGAAGTCGGC-CT  667039

Query  1717   ggc-gctggc-ggcgg-cg-caggcgctgggcgaggc-gggccgggcggtccgcgTGGTG  1771
              ||| | || ||| || || ||||| || ||| | || ||||||||||| |||||||||||
Sbjct  667040  GGCCG-TG-CAGGCCTACGACAAGCTCAG--CGAG-CAGGGCCGCAAGGTCCGCGTGGTG  667094

Query  1772   TCGGTGCCGTGC-CTGGAGCGTTT-CGAGCGGCAGG-CGGCGGACTACC-GCGAGGCGGT  1827
              ||| |||||||| | ||||| |||| | || || ||||| || |||| ||| || ||||||
Sbjct  667095  TCGATGCCGTGCACC--AGCGTCTACGAGCAGCAGGACGA-GTCCTACAAGC-AGTCCGT  667150

Query  1828   GCTGCCG-TCCGCGGTGCGG-GCCCGCGTCGCGGTCGAGATGGCGCGT-CCGGAG-AGCT  1883
              | |||||  || |  || || ||| ||||||  |||||| ||||| | |||| || ||||
Sbjct  667151  GTTGCCGGTG-GAAGT-CGGCGCGCGCGCATCGCCATCGAGGCCGCCCATGCCG-ACTA-CT  667206
```

```
Query  1884   GGCACC-GCTACGTGGGGCTCGACGG-CGCGGTGC-TCGGCATGGAGGGG-TTCGGCGAG   1939
              || ||  |  |||||  ||  ||||||||  |||  |  |  ||||||||  |   |  ||||||||||
Sbjct  667207  GGTACAAG-TACGTCGGTCTCGACGGGCGCA-T-CATCGGCATG-ACCAGCTTCGGCGAG   667262

Query  1940   TCGGCGCCGGCGGCGGAGCTGTTCCG-GCATTTCGGCTTCA   1979
              |||||||||||  |||  ||| |||||  |  |||  ||||||||||
Sbjct  667263  TCGGCGCCGGCCCCGGCGCTGTTC-GAGCACTTCGGCTTCA   667302
```

G.1.2 BLAST result of *tktA2* gene on NCBI shows the best hits against *Pseudomonas aeruginosa* with 78% similarity.

```
 Score = 1024 bits (554), Expect = 0.0
 Identities = 1562/2021 (78%), Gaps = 179/2021 (8%)
 Strand=Plus/Plus


Query  47     GCCAATGCCATCCGCGTC-CTGAGCATGGATGCGGTG--GAACGGGCGAACTCCGGCCAT   103
              |||||||||||||||| | | |||||||||||||| |||  |||  || ||| |||||
Sbjct  665373  GCCAATGCCATCCGTG-CACTGAGCATGGATGCCGTGCAGAA--AGCCAACAGCGGCCA-   665428

Query  104    CCC-GGCGCCCCCATGGGCATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGC-   161
              ||| |||||||||| ||||||||||||||||||||| || |||||| || |  ||
Sbjct  665429  CCCGGGCGCCCCCGATGGGCATGGCCGATATCGCCGAGGTCCTCTGGCGCGACTACATGCA   665488

Query  162    GCCACAACCCGGGCAACCC-CCGCTGGCCCGACCGCGACCGCTTCGTGCTCTCCAACGGG   220
              | ||||||||| ||||||| |  | |||  || ||||||||||||||||||||| ||||||||
Sbjct  665489  G-CACAACCCGAGCAACCCGCAG-TGGGCCAACCGCGACCGCTTCGTGCTGTCCAACGGC   665546

Query  221    CACGGCTCGATGCTGCTCTACAGCCTGCTGCACCTCTCCGGCTACGA-CTTGCCGGTCGA   279
              |||||||||||||||| ||||| |||||||||||| |||| ||||| |  |||
Sbjct  665547  CACGGCTCGATGCTGATCTACTCCCTGCTGCACCTCACCGGGTACGACCTCGGC-ATCGA   665605

Query  280    GGAGCTCAAGCGCTTCCGCCAGTTGC-ATTCGCGCACTCCCGGG-CACCCGGAGTACGGC   337
              ||| || ||| |||||||||||||| | | | ||||||||| |||||| |||||||| ||||||
Sbjct  665606  GGACCTGAAGAACTTCCGCCAGCT-CAACTCGCGCAC-CCCGGGTCACCCGGAATACGGC   665663

Query  338    TACA-CGCCGGGGGTGGAGACCACCACCGGTCCGCTCGGCCAGGGGC-TGGCCAACGCCG   395
              ||||  ||| |||| || |||||||||||||||||||||||||  |||| | ||||| || |
Sbjct  665664  TACACCGCC-GGCGTCGAGACCACCACCGGTCCGCTCGGCCA-GGGCATCGCCAATGCGG   665721

Query  396    TCGGCATGGCCCTGGCCGAG-CGGCTGCTGGCGGCGGACTTCAACCGC-CCGGGACAGGC   453
              | ||||||||| ||||| ||| || | ||||| ||  | ||||||||| | || || ||
Sbjct  665722  TGGGCATGGCGCTGGCGGAGAAGG-TCCTGGCCGCCCAGTTCAACCGCGAC-GGCCACGC   665779

Query  454    GATCGTCGACCACCGCACCTACGCCTTCGTCGGCGACGGCTGCCTGATGGAGGGCATCTC   513
              | | ||||||||| |||||||||||||| |||||||||  ||| ||| ||||| ||||| ||
Sbjct  665780  GGTGGTCGACCACTACACCTACGCCTTCCTCGGCGACGGTTGCATGATGGAAGGCATTTC   665839

Query  514    CCACGAGGTCTG-CTCGCTGGCCGGCACCCTGGGACTCG-GCAAGCTGATCGTGTTCTAC   571
              ||| ||||||  | | |||||||||||||||||||| | || | || ||||||||| ||||||
Sbjct  665840  CCATGAGGTC-GCCTCGCTGGCCGGCACCCTGCGCCT-GAACAAGCTGATCGCCTTCTAC   665897

Query  572    GACGACAACGGCATCTCCATCGACGGTGAGGTCG-CGGGCTGGTTCAGCGACGACACGCC   630
              |||||||||||||||||||||||||||  ||||||  ||| |||||||| ||||||| ||
Sbjct  665898  GACGACAACGGCATCTCCATCGACGGCGAGGTCCACGG-CTGGTTCACCGACGACACCCC   665956

Query  631    GGCGCGCTTCCGCG-CCTACGGCTGGCAGGTCG-TC-G-AGGCGGTCGACGGCCACGACG   686
              | ||||||| |||| |||||||||||||||| || |  |  | ||| ||||| || |||||
Sbjct  665957  GAAGCGCTTC-GAGGCCTACGGCTGGCAGGT-GATCCGCAAT-G-TCGACGGGCATGACG   666012

Query  687    CC-AGCGCGCTCGAGGCGGCGC-TGG---CGGCG-GCCAAG-GCCGATACGGAGCGGCCG   739
              || | || |  || || ||  | |   | | ||| || || || |  | | ||| | |
Sbjct  666013  CCGA-CGAGATCAAGACCGC-CATCGATACCGCGCGC-AAGAGC-GAC-C--AGC--C-G   666062

Query  740    ACGCTGATCTGCTGCAAGACCACCATCGGCTTCGGCGCGCCGACCAAGGCCGG-CA-GCC   797
              || ||||||||||||||||||| ||||||||||||| ||||||| |||| ||| | | | ||
Sbjct  666063  ACCCTGATCTGCTGCAAGACCGTGATCGGCTTCGGCTCGCCGAACAAG-CAGGGCAAGGA   666121

Query  798    ACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGATCGCCGCGGCCCGCGAGGCGCTGG   857
              | | || ||||||||| |||||||||||||||| ||| |||||||||||||| |||||| |||||||
Sbjct  666122  A-GAGTGCCACGGCGCGCCGCTGGGCGCCGACGAGATCGCCGCGACCCGCGCCGCGCTGG   666180

Query  858    ACTGGCCGCACGCGCCC-TTCGAGGTGCCCG-CGCAGATC---GCCAGGGCCTGGGACGC   912
              |||| ||||| |||||| | || ||||| |||| ||||||||  ||| |  ||||||||
Sbjct  666181  GCTGGGAGCACG-GTCCGTTCGAGAT-CCCGGCGCAGATCTACGCC-GAG--TGGGACGC   666235
```

105

```
Query  913   GC-GCGAG-CGCGGGGCGGCGCTGGA-GCGGGACTGGCGACAA-CGCTTCGACGACTACG  968
             | | ||   | ||| || ||  ||| || | ||  |||  |||| || |||| || ||||
Sbjct  666236 -CAGGGAAAC-CGGCGCCGCCCAGGAAGCCG-AGTGGA-ACAAGCGTTTCGCCGCCTAC-  666290

Query  969   C-G-CGGGCCTACCCGGCGCTGGCCGCCGAGCTGGAGCGGCGCCTGGCCGGCGAGCTGCC  1026
             | | | || | |||| |||||||||||| ||| ||| |||||| |||||||||||
Sbjct  666291 CAGGCCGCCC-ATCCGGAACTGGCCGCCGAACTGCTGCGCCGCCTGAAGGGCGAGCTGCC  666349

Query  1027  GGCCGACTGGggcgcgc-ggcgcgac-gcgct--gatcgccg--g-ctgccagcaggcgg  1079
             ||||||||| | || |  ||| || | || || | ||||| | | |||| || | ||
Sbjct  666350 GGCCGACTTCGCCGAGAAGGC-CG-CTGC-CTACG-TCGCCGATGTC-GCCAACAAG-GG  666403

Query  1080  cgcg-cgagctggcgACCCGCAAGAGCTC-GC-AGCAGGTC-CTGGAAGCGC-TG-GGCC  1133
             || | | | | | ||||||| || | | | ||  | |  || | ||| |  |  ||||
Sbjct  666404 CGAGACCATC-GCC-AGCCGCAAG-GC-CAGCCAGAACG-CGCTG-AA-CGCCTTCGGCC  666456

Query  1134  CGCTGCTGCCGGAACTGCTCGGCGGCTCGGCGGACCTGGC-GCCCTCGAACCTGACCC-G  1191
             |||||||||||| |||||||||||| || || ||||||| | |||||||||||||||| |
Sbjct  666457 CGCTGCTGCCGGAGCTGCTCGGCGGTTCCGCCGACCTGGCCGGC-TCCAACCTGACCCTG  666515

Query  1192  CTGGTCCGGCTCCC-GCTCGGTCGGCGGC-GAGG-CGCCCGGGGGCAACTACATCCACTA  1248
             | |  |||| |  | || |    | |||| |   | ||||| ||||||||| | || || |
Sbjct  666516 -TGGAAGGGCTGTAAGGGCG-TCAGCG-CTGACGACGCC-GCCGGCAACTACGTGTTCTA  666571

Query  1249  CGGGGTGCGCGAGTTCGGCATGAGCGCGATGATGAACGG-GCTGGCGCTGCACGGCGGTT  1307
             |||  ||||||||||||||||||||||||||| ||||||||| ||||||| | | || |||||||||||||
Sbjct  666572 CGGCGTGCGCGAGTTCGGCATGAGCGCGATCATGAACGGCG-TCGCCCTGCACGGCGGTT  666630

Query  1308  TCATTCCCTACGGCG-GCACCTTCCTGATGTTCATGGAATACGCCCGCAACGCCGTGCGC  1366
             ||||||||||||||| | |||||||||||| || |||||| ||||| |||||||||||||||||||||
Sbjct  666631 TCATTCCCTACGGCGCG-ACCTTCCTGATCTTCATGGAGTACGCGCGCAACGCCGTGCGC  666689

Query  1367  ATGGCCGCGCTGATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCCTG  1426
             ||| |||| ||||||||||||| ||||| | | | ||| | ||||||  |||| ||||||||||
Sbjct  666690 ATGTCCGCCCTGATGAAGCAGCGCGTGCTCTACGTGTTCACCCATGACTCCATCGGCCTC  666749

Query  1427  GGCGAGGACGGCCCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCGCCAGACGCCG  1486
             |||||||||||||||||||| ||||||||| |||| |  | || |||||||||||||| ||| ||||
Sbjct  666750 GGCGAGGACGGCCCGACCCACCAGCCGATCGAGCAACTGGCCAGCCTGCGCCTGACCCCG  666809

Query  1487  AACCTGGAAACCTGGCGGCCCTGC-GACGAGA-CGGAGACCGCGGTGGCCTGGT-GCGCG  1543
             ||||||||| |||||||| | |||| ||| || || |  || || | |||||||||||| ||  |
Sbjct  666810 AACCTGGACACCTGGCG-CCCGGCCGATGCGGTCG-AGTCGGCGGTGGCCTGGAAGCATG  666867

Query  1544  GCGCTCGAACG-GCGGGACGGCCCGTCGGCGCTGGTGCTG-Tcgcggcagaacctgccgg  1601
             |   |||| || ||  |  |||| ||||| ||||| | | | | ||||||||||||||||
Sbjct  666868 CCA-TCGAGCGTGCCG-ACGGTCCGTCCGCGCTGAT-CTTCTCCCGCCAGAACCTGCCG-  666923

Query  1602  ccgc-ggcgcgcgacgccgggcagcgcg-cggccatcgagcgcggcgccTATGTGCTGCT  1659
             |   | |||||||||| ||  ||| | | | | | ||||| |||||||| ||| || |||
Sbjct  666924 CATCAGGCGCGCGACGTCGCCCAG-GTGACCGACATCGCGCGCGGCGGCTACGTACTGAA  666982

Query  1660  CGACTGC-CCGG-GCACGCCGGAGCTGATCCTgctggcgtccggctccgaggt-ggcgct  1716
             |||||| |||| | | | |||||| |||||||| | ||  |||| || || ||| ||
Sbjct  666983 GGACTGCGCCGGCG-A-GCCGGAACTGATCCTGATCGCCACCGGTTCGGAAGTCGGC-CT  667039

Query  1717  ggc-gctggc-ggcgg-cg-caggcgctgggcgaggc-gggccgggcggtccgcgTGGTG  1771
             ||| | || | || ||  || | | || ||||| | |  |||| | |||| ||||||||||||||
Sbjct  667040 GGCCG-TG-CAGGCCTACGACAAGCTCAG--CGAG-CAGGGCCGCAAGGTCCGCGTGGTG  667094

Query  1772  TCGGTGCCGTGC-CTGGAGCGTTT-CGAGCGGCAGG-CGGCGGACTACC-GCGAGGCGGT  1827
             ||| |||||||| |     ||||| | ||||| ||||| ||  |   |||| || || ||
Sbjct  667095 TCGATGCCGTGCACC--AGCGTCTACGAGCAGCAGGACGA-GTCCTACAAGC-AGTCCGT  667150

Query  1828  GCTGCCG-TCCGCGGTGCGG-GCCCGCGTCGCGGTCGAGATGGCGCGT-CCGGAG-AGCT  1883
             | ||||| |    || || ||  | ||| ||| || || | ||||  || | |  ||
Sbjct  667151 GTTGCCGGTG-GAAGT-CGGCGCGCGCGCATCGCCATCGAGGCCGCCCATGCCG-ACTA-CT  667206

Query  1884  GGCACC-GCTACGTGGGTCTCGACGG-CGCGGTGC-TCGGCATGGAGGGG-TTCGGCGAG  1939
             || ||  | ||||| ||||||||||| ||| | | | |||||||| |    | ||||||||
Sbjct  667207 GGTACAAG-TACGTCGGTCTCGACGGGCGCA-T-CATCGGCATG-ACCAGCTTCGGCGAG  667262

Query  1940  TCGGCGCCGGCGGCGGAGCTGTTCCG-GCATTTCGGCTTCA  1979
             ||||||||||||   ||| ||||||| | ||| ||||||||||
Sbjct  667263 TCGGCGCCGGCCCCGGCGCTGTTC-GAGCACTTCGGCTTCA  667302
```

106

G.1.3 BLAST result of *tktA3* gene on NCBI shows the best hits against *Pseudomonas aeruginosa* with 85% similarity.

```
 Score = 2023 bits (1095),  Expect = 0.0
 Identities = 1739/2040 (86%), Gaps = 84/2040 (4%)
 Strand=Plus/Plus


Query  1       ATGCCCAGCCGTCGTGAGCGAGCCAATGCCATTCGTGCCCTGAGTATGGATGCCGTGCAG  60
               ||||||||||||||||||||||||||||||||||| ||||| ||||| ||||||||||||
Sbjct  665352  ATGCCCAGCCGTCGTGAGCGAGCCAATGCCATCCGTGCACTGAGCATGGATGCCGTGCAG  665411

Query  61      AAAGCCAACAGTGGCCATCCTGGCGCCCCCATGGGCATGGCGGATATCGCCGAAGTGCTG  120
               ||||||||||| ||||| || |||||||| ||||||||||| ||||||||| || ||
Sbjct  665412  AAAGCCAACAGCGGCCACCCGGGCGCCCCCGATGGGCATGGCCGATATCGCCGAGGTCCTC  665471

Query  121     TGGCACGACTATC-T-CAGGCACAACCCG-GCCAATCCGCAATGGGCCGACCGCGACCGC  177
               |||| ||||||| | | || |||||||||| | ||| ||||| |||||| ||||||||||
Sbjct  665472  TGGCGCGACTA-CATGCA-GCACAACCCGAG-CAACCCGCAGTGGGCCAACCGCGACCGC  665528

Query  178     TTCGTGCTCTCCAACGGCCATGGCTCGATGCTGGTCTATTCGCTGCTGCATCTGACCGGC  237
               |||||||| ||||||||||| |||||||||||| |||| || || |||||||| |||||
Sbjct  665529  TTCGTGCTGTCCAACGGCCACGGCTCGATGCTGATCTACTCCCTGCTGCACCTCACCGGG  665588

Query  238     TACGACCTCTCG-ATCGACGACCTGAAGAACTTCCGCCAGTTCGGCAGC-CGCACCCCCG  295
               ||||||||| |  | ||||| |||||||||||||||||||||||| || |  |  | ||||||||| |
Sbjct  665589  TACGACCTC-GGCATCGAGGACCTGAAGAACTTCCGCCAGCTCAAC-TCGCGCACCCCGG  665646

Query  296     GCCATCCC-GAATACGGCTATACCCCCGGCGTCGAGACCACCACCGGCCCGCTCGGCCAG  354
               | || ||| ||||||||||| ||| ||||||||||||||||||||||||||| ||||||||||||||
Sbjct  665647  GTCA-CCCGGAATACGGCTACACCGCCGGCGTCGAGACCACCACCGGTCCGCTCGGCCAG  665705

Query  355     GGCCTGGCCAACGCCGTCGGCATGGCCG-TGGCGGAGAAGGTGCTGGCCGCGCAGTTCAA  413
               ||| | ||||| || || ||||||| || | ||||||||||||| ||||||| |||||||| ||||||
Sbjct  665706  GGCATCGCCAATGCGGTGGGCATGG-CGCTGGCGGAGAAGGTCCTGGCCGCCCAGTTCAA  665764

Query  414     CCGGC-CCGGGCAGCGC-ATCGTCGATCATTACACCTACGTGTTCCTCGGCGACGGCTGC  471
               || || ||| || ||| | ||| |||| || ||||||||| |||||||||||||| |||
Sbjct  665765  CC-GCGACGGCCA-CGCGGTGGTCGACCACTACACCTACGCCTTCCTCGGCGACGGTTGC  665822

Query  472     CTGATGGAAGGCATCTCCCACGAAGTCTG-CTCGCTGGCCGGCACCCTGGGGCTGAACAA  530
               ||||||||||||||| ||| || || | || |||||||||||||||||| ||||||||||
Sbjct  665823  ATGATGGAAGGCATTTCCCATGAGGTC-GCCTCGCTGGCCGGCACCCTGCGCCTGAACAA  665881

Query  531     GCTGGTCGCCTTCTACGACGACAACGGCATCTCCATCGACGGCGAGGTCGAGGGCTGGTT  590
               |||| |||||||||||||||||||||||||||||||||||||||||||| | ||||||||
Sbjct  665882  GCTGATCGCCTTCTACGACGACAACGGCATCTCCATCGACGGCGAGGTCCACGGCTGGTT  665941

Query  591     CACCGACGACACCCCGGCGCGCTTCGAGGCCTACGGCTGGCTGGTGATCCGCAACGTCGA  650
               |||||||||||||||| ||||||||||||||||||||||||| ||||||||||| |||||
Sbjct  665942  CACCGACGACACCCCGAAGCGCTTCGAGGCCTACGGCTGGCAGGTGATCCGCAATGTCGA  666001

Query  651     CGGCCACGACGCCGACGAGATCCGGACCGCGATCGAGACCGCGCGCAAGAGCGAGC-GCC  709
               ||| || ||||||||||||||| ||||| ||||| ||||||||||||||||||||| | | |
Sbjct  666002  CGGGCATGACGCCGACGAGATCAAGACCGCCATCGATACCGCGCGCAAGAGCGACCAG-C  666060

Query  710     CGGTCCTGATCTGCTGCAAGACGGTGATCGGTTTCGGTTCGCCGAACAAGCAGGGCAAGG  769
               || ||||||||||||||||||| ||||| |||||| |||| ||||||||||||||||||||
Sbjct  666061  CGACCCTGATCTGCTGCAAGACCGTGATCGGCTTCGGCTCGCCGAACAAGCAGGGCAAGG  666120

Query  770     AAGCCTGCCACGGCGCCG-CGCTGGGGGCCGAGGAAATCGCC-CTGACCCGCGACGCGCT  827
               ||| |||||||||| || ||||||| ||||||| || || |||||| | ||||||||| ||||||
Sbjct  666121  AAGAGTGCCACGGC-GCCGCTGGGCGCCGACGAGATCGCCGC-GACCCGCGCCGCGCT  666178

Query  828     GGGCTGGAAGCACGCG-CCGTTCGAGATTCCCGCCG-AGATCTACGCCG-GCTGGGACGC  884
               |||||||| ||||| | |||||||||| |||||| |||||||||||||| | ||||||||
Sbjct  666179  GGGCTGGGAGCACG-GTCCGTTCGAGA-TCCCGGCGCAGATCTACGCCGAG-TGGGACGC  666235

Query  885     CAGGGCCGCCGGCGCCGCCC-GCG-AGACCGAATGGAACCGGCGCTTCGCCGCCTACCAG  942
               ||||| ||||||||||||| | | || |||| ||||| ||| ||| ||||||||||||||
Sbjct  666236  CAGGGAAACCGGCGCCGCCCAG-GAAG-CCGAGTGGAACAAGCGTTTCGCCGCCTACCAG  666293

Query  943     GCCG-AGTTCCCGGAACTGGCCGCGGAGTTCCTGCGCCG-CAGCCGTGGCGAACTGCCCG  1000
               ||||    | ||||||||||||||||| || | ||||||||| | | | ||||| ||||| |
Sbjct  666294  GCCGCCCAT-CCGGAACTGGCCGCCGAACTGCTGCGCCGCCTGAAG-GGCGAGCTGCCGG  666351
```

```
Query  1001    CCGATTTCGCCGCGAAGGCCGGC-GAATACATCGCCGAAATCGCCGCCAAGGGGGAAACC      1059
               |||| ||||||| |||||||| || |    ||| ||||||| ||||| ||||||| || |||
Sbjct  666352  CCGACTTCGCCGAGAAGGCC-GCTGCCTACGTCGCCGATGTCGCCAACAAGGGCGAGACC      666410

Query  1060    ATCGCCAGCCGCAAGGCCAGCCAGAACGCGCTGAACGCCTTCGGCCCGCTGCTGCCGGAG      1119
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  666411  ATCGCCAGCCGCAAGGCCAGCCAGAACGCGCTGAACGCCTTCGGCCCGCTGCTGCCGGAG      666470

Query  1120    TTCCTCGGCGGCTCGGCGGACCTGGCCGGTTCCAACCTGACCCTGTGGAAGGGCTGCAAG      1179
               | ||||||||| || || |||||||||||  ||||||||||||||||||||||||||  |||
Sbjct  666471  CTGCTCGGCGGTTCCGCCGACCTGGCCGGCTCCAACCTGACCCTGTGGAAGGGCTGTAAG      666530

Query  1180    G-CGCTC-GCGGCGGACGACGCCTCCGGTAACTACCTGCACTACGGCGTGCGCGAGTTCG      1237
               | || ||| | |||||||| |||| ||||  |||||| || |||||||||||||||||||||
Sbjct  666531  GGCG-TCAGCG-CTGACGACGCCGCCGGCAACTACGTGTTCTACGGCGTGCGCGAGTTCG      666588

Query  1238    GCATGAGCGCGATCATGAACGGTATCGCTCTGCACGGCGGCTTCATTCCCTACGGCGCCA      1297
               |||||||||||||||||||||||   |||| ||||||||||| |||||||||||||||| |
Sbjct  666589  GCATGAGCGCGATCATGAACGGCGTCGCCCTGCACGGCGGTTTCATTCCCTACGGCGCGA      666648

Query  1298    CCTTCCTGGTGTTCATGGAGTACGC-CTGCAACGCCGTGCGCATGGCCGCGCTGATGAAG      1356
               |||||| |  |||||||||||||||| ||||||||||||||||||||  | | |||||||||
Sbjct  666649  CCTTCCTGATCTTCATGGAGTACGCGC-GCAACGCCGTGCGCATGTCCGCCCTGATGAAG      666707

Query  1357    CAGCGCGTGCTGTTCGTCTACACCCACGACTCCATCGGCCTCGGCGAGGACGGCCCGACC      1416
               |||||||||||| | ||| |  |||||| |||||||||||||||||||||||||||||||||
Sbjct  666708  CAGCGCGTGCTCTACGTGTTCACCCATGACTCCATCGGCCTCGGCGAGGACGGCCCGACC      666767

Query  1417    CACCAGCCGGTCGAGCAACTGGCCAGCCTGCGCA-GCACGCCGAACCTCGACACCTGGCG      1475
               ||||||||| | ||||||||||||||||||||||| | || ||||||| |||| |||||||||
Sbjct  666768  CACCAGCCGATCGAGCAACTGGCCAGCCTGCGCCTG-ACCCCGAACCTGGACACCTGGCG      666826

Query  1476    CCCGTGC-GACGCGGTGGAGTCGGCGGTGGCCTGGAA--ATGCGCCATCGAGCG--CCAG      1530
               |||| || || |||| || |||||||||||||||||||  |||| | |||||||||  || |
Sbjct  666827  CCCG-GCCGATGCGGTCGAGTCGGCGGTGGCCTGGAAGCATGC-C-ATCGAGCGTGCC-G      666882

Query  1531    GACGGCCCGAGCACCCTGGTGTTCAGCCGCCAGAACCTGCCGCACCAGGCGCGCGATGCC      1590
               |||| |||  | | ||| | |||  |||||||||||||||||||| |||||||||||| | |
Sbjct  666883  -ACGGTCCGTCCGCGCTGATCTTCTCCCGCCAGAACCTGCCGCATCAGGCGCGCGACGTC      666941

Query  1591    CGGCAAC-TGGCCGATATCGCCCGCGGCGGCTACGTGCTCC-GCGATTGTGCCGGC-ACC      1647
               | | |   || |||| ||||| ||||| |||||||||||||| ||    | || || |||| | |
Sbjct  666942  -GCCCAGGTGACCGACATCGCGCGCGGCGGCTACGTACTGAAG-GACTGCGCCGGCGAGC      666999

Query  1648    CCGGAGCTGATCCTCATCGCCACCGGTTCGGAGGTCGCCCTGGCGGTGCAGGCCCACGAG      1707
               | ||| |||||||| |||||||||||||||||||| |||| ||||||| |||||| ||||
Sbjct  667000  C-GGAACTGATCCTGATCGCCACCGGTTCGGAAGTCGGCCTGGCCGTGCAGGCCTACGAC      667058

Query  1708    ACGCTGAGCGGGCAGGGCCGCCGGGTGCGAGTGGTTTCCATGCCCT-CGACCAGCGTGTT      1766
               | ||| |||| |||||||||||||||  ||| || || || ||| |||  | | |||||||||| |
Sbjct  667059  AAGCTCAGCGAGCAGGGCCGCAAGGTCCGCGTGGTGTCGATGCCGTGC-ACCAGCGTCTA      667117

Query  1767    CGATC-GCCAGGACGCCG-CCTGGAAACAGGCCGTGCTGCCG-TCCGGG-GTCGGGGCGC      1822
               ||| | | || |||||| | ||| || ||| ||||| ||||| ||  || ||||| ||||||
Sbjct  667118  CGAGCAGC-AGGACGA-GTCCTACAAGCAGTCCGTGTTGCCGGT--GGAAGTCGGCGCGC      667173

Query  1823    GGATCGCCATCGAGGCCGCCCACGCCGACTTCTGGTACAAGTACGTCGGGCTGG-CTGGA      1881
               | |||||||||||||||||||| || |||| ||||||||||||||||||    || | | |
Sbjct  667174  GCATCGCCATCGAGGCCGCCCATGCCGACTACTGGTACAAGTACGTCGGTCTCGACGGG-      667232

Query  1882    CGCATCGTCGGCATGACCCGCTTCGGCGAGTCCGCGCCGGCCGGCG-CGCTCTTCGAGCA      1940
               |||||| |||||||||||| ||||||||||||| ||||||||||  | || |||| |||||||
Sbjct  667233  CGCATCATCGGCATGACCAGCTTCGGCGAGTCGGCGCCGGCCC-CGGCGCTGTTCGAGCA      667291

Query  1941    TTTCGGCTTCACCCTCGACAACCTGCTGGCCAGCG-CCGGGGAACTGCTCGGG-GACTGA      1998
               |||||||||||||||| |||||| | ||||||  | | ||| |||||||||| || ||||||
Sbjct  667292  CTTCGGCTTCACCCTGGACAACGTCCTGGCG-GTGGCCGAGGAACTGCT-GGAAGACTGA      667349
```

108

G.2 The multiple alignment of the three transketolase shows a 98% similarity of *tktA1* to *tktA2* while *tktA3* is 75% similar to both *tktA1* and *tktA2*.

```
SeqA Name    Len(nt)  SeqB Name    Len(nt)  Score
==================================================
1    tktA1   2176     2    tktA2   2113     98
1    tktA1   2176     3    tktA3   1998     75
2    tktA2   2113     3    tktA3   1998     75
==================================================
```

```
tktA1      GATAACGCATATGACCCAGAACCTGAAACCGGGCCGACGCGCCCTGGCCAATGCCATCCG 60
tktA2      GATAACGCATATGACCCAGAACCTGAAACCGGGCCGACGCGCCCTGGCCAATGCCATCCG 60
tktA3                          ATGCCCAGCCGTCGTGAGCGAGCCAATGCCATTCG 35

tktA1      CGTCCTGAGCATGGATGCGGTGGAACGGGCGAACTCCGGCCATCCCGGCGCCCCCATGGG 120
tktA2      CGTCCTGAGCATGGATGCGGTGGAACGGGCGAACTCCGGCCATCCCGGCGCCCCCATGGG 120
tktA3      TGCCCTGAGTATGGATGCCGTGCAGAAAGCCAACAGTGGCCATCCTGGCGCCCCCATGGG 95

tktA1      CATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGCGCCACAACCCGGGCAACCC 180
tktA2      CATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGCGCCACAACCCGGGCAACCC 180
tktA3      CATGGCGGATATCGCCGAAGTGCTGTGGCACGACTATCTCAGGCACAACCCGGCCAATCC 155

tktA1      CCGCTGGCCCGACCGCGACCGCTTCGTGCTCTCCAACGGGCACGGCTCGATGCTGCTCTA 240
tktA2      CCGCTGGCCCGACCGCGACCGCTTCGTGCTCTCCAACGGGCACGGCTCGATGCTGCTCTA 240
tktA3      GCAATGGGCCGACCGCGACCGCTTCGTGCTCTCCAACGGCCATGGCTCGATGCTGGTCTA 215

tktA1      CAGCCTGCTGCACCTCTCCGGCTACGACTTGCCGGTCGAGGAGCTCAAGCGCTTCCGCCA 300
tktA2      CAGCCTGCTGCACCTCTCCGGCTACGACTTGCCGGTCGAGGAGCTCAAGCGCTTCCGCCA 300
tktA3      TTCGCTGCTGCATCTGACCGGCTACGACCTCTCGATCGACGACCTGAAGAACTTCCGCCA 275

tktA1      GTTGCATTCGCGCACTCCCGGGCACCCGGAGTACGGCTACACGCCGGGGGTGGAGACCAC 360
tktA2      GTTGCATTCGCGCACTCCCGGGCACCCGGAGTACGGCTACACGCCGGGGGTGGAGACCAC 360
tktA3      GTTCGGCAGCCGCACCCCCGGCCATCCCGAATACGGCTATACCCCCGGCGTCGAGACCAC 335

tktA1      CACCGGTCCGCTCGGCCAGGGGCTGGCCAACGCCGTCGGCATGGCCCTGGCCGAGCGGCT 420
tktA2      CACCGGTCCGCTCGGCCAGGGGCTGGCCAACGCCGTCGGCATGGCCCTGGCCGAGCGGCT 420
tktA3      CACCGGCCCGCTCGGCCAGGGCCTGGCCAACGCCGTCGGCATGGCCGTGGCGGAGAAGGT 395

tktA1      GCTGGCGGCGGACTTCAACCGCCCGGGACAGGCGATCGTCGACCACCGCACCTACGCCTT 480
tktA2      GCTGGCGGCGGACTTCAACCGCCCGGGACAGGCGATCGTCGACCACCGCACCTACGCCTT 480
tktA3      GCTGGCCGCGCAGTTCAACCGGCCCGGGCAGCGCATCGTCGATCATTACACCTACGTGTT 455

tktA1      CGTCGGCGACGGCTGCCTGATGGAGGGCATCTCCCACGAGGTCTGCTCGCTGGCCGGCAC 540
tktA2      CGTCGGCGACGGCTGCCTGATGGAGGGCATCTCCCACGAGGTCTGCTCGCTGGCCGGCAC 540
tktA3      CCTCGGCGACGGCTGCCTGATGGAAGGCATCTCCCACGAAGTCTGCTCGCTGGCCGGCAC 515

tktA1      CCTGGGACTCGGCAAGCTGATCGTGTTCTACGACGACAACGGCATCTCCATCGACGGTGA 600
tktA2      CCTGGGACTCGGCAAGCTGATCGTGTTCTACGACGACAACGGCATCTCCATCGACGGTGA 600
tktA3      CCTGGGGGCTGAACAAGCTGGTCGCCTTCTACGACGACAACGGCATCTCCATCGACGGCGA 575

tktA1      GGTCGCGGGCTGGTTCAGCGACGACACGCCGGCGCGCTTCCGCGCCTACGGCTGGCAGGT 660
tktA2      GGTCGCGGGCTGGTTCAGCGACGACACGCCGGCGCGCTTCCGCGCCTACGGCTGGCAGGT 660
tktA3      GGTCGAGGGCTGGTTCACCGACGACACCCCGGCGCGCTTCGAGGCCTACGGCTGGCTGGT 635

tktA1      CGTCGAGGCGGTCGACGGCCACGACGCCAGCGCGCTCGAGGCGGCGCTGGCGGCGGCCAA 720
tktA2      CGTCGAGGCGGTCGACGGCCACGACGCCAGCGCGCTCGAGGCGGCGCTGGCGGCGGCCAA 720
tktA3      GATCCGCAACGTCGACGGCCACGACGCCGACGAGATCCGGACCGCGATCGAGACCGC--- 692

tktA1      GGCCGATACGGAGCGGCCGACGCTGATCTGCTGCAAGACCACCATCGGCTTCGGCGCGCC 780
tktA2      GGCCGATACGGAGCGGCCGACGCTGATCTGCTGCAAGACCACCATCGGCTTCGGCGCGCC 780
tktA3      GCGCAAGAGCGAGCGCCCGGTCCTGATCTGCTGCAAGACGGTGATCGGTTTCGGTTCGCC 752

tktA1      GACCAAGGCCGGCAGCCACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGATCGCCGC 840
tktA2      GACCAAGGCCGGCAGCCACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGATCGCCGC 840
tktA3      GAACAAGCAGGGCAAGGAAGCCTGCCACGGCGCCGCGCTGGGGGCCGAGGAAATCGCCCT 812
```

```
tktA1    GGCCCGCGAGGCGCTGGACTGGCCGCACGCGCCCTTCGAGGTGCCCGCGCAGATCGCCAG 900
tktA2    GGCCCGCGAGGCGCTGGACTGGCCGCACGCGCCCTTCGAGGTGCCCGCGCAGATCGCCAG 900
tktA3    GACCCGCGACGCGCTGGGCTGGAAGCACGCGCCGTTCGAGATTCCCGCCGAGATCTACGC 872

tktA1    GGCCTGGGACGCGCGCGAGCGCGGGGCGGCGCTGGAGCGGGACTGGCGACAACGCTTCGA 960
tktA2    GGCCTGGGACGCGCGCGAGCGCGGGGCGGCGCTGGAGCGGGACTGGCGACAACGCTTCGA 960
tktA3    CGGCTGGGACGCCAGGGCCGCCGGCGCCGCCCGCGAGACCGAATGGAACCGGCGCTTCGC 932

tktA1    CGACTACGCGCGGGCCTACCCGGCGCTGGCCGCCGAACTGGAGCGGCGCCTGGCCGGCGA 1020
tktA2    CGACTACGCGCGGGCCTACCCGGCGCTGGCCGCCGAGCTGGAGCGGCGCCTGGCCGGCGA 1020
tktA3    CGCCTACCAGGCCGAGTTCCCGGAACTGGCCGCGGAGTTCCTGCGCCGCAGCCGTGGCGA 992

tktA1    GCTGCCGGCCGACTGGGGCGCGCGGCGCGACGCGCTGATCGCCGG--CTGCCAGCAGGCG 1078
tktA2    GCTGCCGGCCGACTGGGGCGCGCGGCGCGACGCGCTGATCGCCGG--CTGCCAGCAGGCG 1078
tktA3    ACTGCCCGCCGATTTCGCCGCGAAGGCCGGCGAATACATCGCCGAAATCGCCGCCAAGGG 1052

tktA1    GCGCGCGAGCTGGCGACCCGCAAGAGCTCGCAGCAGGTCCTGGAAGCGCTGGGCCCGCTG 1138
tktA2    GCGCGCGAGCTGGCGACCCGCAAGAGCTCGCAGCAGGTCCTGGAAGCGCTGGGCCCGCTG 1138
tktA3    GGAAACCATC--GCCAGCCGCAAGGCCAGCCAGAACGCGCTGAACGCCTTCGGCCCGCTG 1110

tktA1    CTGCCGGAGCTGCTCGGCGGCTCGGCGGACCTGGCGCCCTCGAACCTGACCCGCTGGTCC 1198
tktA2    CTGCCGGAACTGCTCGGCGGCTCGGCGGACCTGGCGCCCTCGAACCTGACCCGCTGGTCC 1198
tktA3    CTGCCGGAGTTCCTCGGCGGCTCGGCGGACCTGGCCGGTTCCAACCTGACCCTGTGGAAG 1170

tktA1    GGCTCCCGCTCGGTCGGCGGCGAGGCGCCCGGGGGCAACTACATCCACTACGGGGTGCGC 1258
tktA2    GGCTCCCGCTCGGTCGGCGGCGAGGCGCCCGGGGGCAACTACATCCACTACGGGGTGCGC 1258
tktA3    GGCTGCAAGGCGCTCGCGGCGGACGACGCCTCCGGTAACTACCTGCACTACGGCGTGCGC 1230

tktA1    GAGTTCGGCATGAGCGCGATGATGAACGGGCTGGCGCTGCACGGCGGTTTCATCCCCTAC 1318
tktA2    GAGTTCGGCATGAGCGCGATGATGAACGGGCTGGCGCTGCACGGCGGTTTCATTCCCTAC 1318
tktA3    GAGTTCGGCATGAGCGCGATCATGAACGGTATCGCTCTGCACGGCGGCTTCATTCCCTAC 1290

tktA1    GGCGGCACCTTCCTGATGTTCATGGAATACGCCCGCAACGCCGTGCGCATGGCCGCGCTG 1378
tktA2    GGCGGCACCTTCCTGATGTTCATGGAATACGCCCGCAACGCCGTGCGCATGGCCGCGCTG 1378
tktA3    GGCGCCACCTTCCTGGTGTTCATGGAGTACGCCTGCAACGCCGTGCGCATGGCCGCGCTG 1350

tktA1    ATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCCTGGGCGAGGACGGC 1438
tktA2    ATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCCTGGGCGAGGACGGC 1438
tktA3    ATGAAGCAGCGCGTGCTGTTCGTCTACACCCACGACTCCATCGGCCTCGGCGAGGACGGC 1410

tktA1    CCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCGCCAGACGCCGAACCTGGAAACC 1498
tktA2    CCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCGCCAGACGCCGAACCTGGAAACC 1498
tktA3    CCGACCCACCAGCCGGTCGAGCAACTGGCCAGCCTGCGCAGCACGCCGAACCTCGACACC 1470

tktA1    TGGCGGCCCTGCGACGAGACGGAGACCGCGGTGGCCTGGTGCGCGGCGCTCGAACGGCGG 1558
tktA2    TGGCGGCCCTGCGACGAGACGGAGACCGCGGTGGCCTGGTGCGCGGCGCTCGAACGGCGG 1558
tktA3    TGGCGCCCCGTGCGACGCGGTGGAGTCGGCGGTGGCCTGGAAATGCGCCATCGAGCGCCAG 1530

tktA1    GACGGCCCGTCGGCGCTGGTGCTGTCGCGGCAGAACCTGCCGGCCGCGGCGCGCGACGCC 1618
tktA2    GACGGCCCGTCGGCGCTGGTGCTGTCGCGGCAGAACCTGCCGGCCGCGGCGCGCGACGCC 1618
tktA3    GACGGCCCGAGCACCCTGGTGTTCAGCCGCCAGAACCTGCCGCACCAGGCGCGCGATGCC 1590

tktA1    GGGCAGCGCGCGGCCATCGAGCGCGGCGCCTATGTGCTGCTCGACTGCCCGGGCACGCCG 1678
tktA2    GGGCAGCGCGCGGCCATCGAGCGCGGCGCCTATGTGCTGCTCGACTGCCCGGGCACGCCG 1678
tktA3    CGGCAACTGGCCGATATCGCCCGCGGCGGCTACGTGCTCCGCGATTGTGCCGGCACCCCG 1650

tktA1    GAGCTGATCCTGCTGGCGTCCGGCTCCGAGGTGGCGCTGGCGCTGGCGGCGGCGCAGGCG 1738
tktA2    GAGCTGATCCTGCTGGCGTCCGGCTCCGAGGTGGCGCTGGCGCTGGCGGCGGCGCAGGCG 1738
tktA3    GAGCTGATCCTCATCGCCACCGGTTCGGAGGTCGCCCTGGCGGTGCAGGCCCACGAGACG 1710

tktA1    CTGGGCGAGGCGGGCCGGGCGGTCCGCGTGGTGTCGGTGCCGT-GCCTGGAGCGTTTCGA 1797
tktA2    CTGGGCGAGGCGGGCCGGGCGGTCCGCGTGGTGTCGGTGCCGT-GCCTGGAGCGTTTCGA 1797
tktA3    CTGAGCGGGCAGGGCCGCCGGGTGCGAGTGGTTTCCATGCCCTCGACCAGCGTGTT-CGA 1769

tktA1    GCGGCAGGCGGCGGACTACCGCGAGGCGGTGCTGCCGTCCGCGGTGCGGGCCCGCGTCGC 1857
tktA2    GCGGCAGGCGGCGGACTACCGCGAGGCGGTGCTGCCGTCCGCGGTGCGGGCCCGCGTCGC 1857
tktA3    TCGCCAGGACGCCGCCTGGAAACAGGCCGTGCTGCCGTCCGGGGTCGGGGCGCGGATCGC 1829

tktA1    GGTCGAGATGGCGCGTCCGGAGAGCTGGCACCGCTACGTGGGGCTCGACGGCGCGGTGCT 1917
tktA2    GGTCGAGATGGCGCGTCCGGAGAGCTGGCACCGCTACGTGGGTCTCGACGGCGCGGTGCT 1917
tktA3    CATCGAGGCCGCCCACGCCGACTTCTGGTACAAGTACGTCGGGCTGGCTGGACGCATCGT 1889

tktA1    CGGCATGGAGGGGTTCGGCGAGTCGGCGCCGGCGGCGGAGCTGTTCCGGCATTTCGGCTT 1977
tktA2    CGGCATGGAGGGGTTCGGCGAGTCGGCGCCGGCGGCGGAGCTGTTCCGGCATTTCGGCTT 1977
tktA3    CGGCATGACCCGCTTCGGCGAGTCCGCGCCGGCCGGCGCGCTCTTCGAGCATTTCGGCTT 1949
```

110

```
tktA1        CAGCGTCGAGAACCTGCTCGCCAGGGCGCAGGCCGTGCTGGCCGGGCTGGAGGGGAGGGC 2037
tktA2        CAGCGTCGAGAACCTGCTCGCCAGGGCGCAGGCCGTGCTGGCCGGGCTGGAGGGGAGGGC 2037
tktA3        CACCCTCGACAACCTGCTGGCCAGCGCCGGGGAACTGCTCGG-GGACTGA---------- 1998

tktA1        GTAGCCATGTATCTGGGCGTCGATTGCGGGACCCAGGGGACCAAACCAGCCGGGTAACAC 2097
tktA2        GTAGCCATGTATCTGGGCGTCGATTGCGGGACCCAGGGGACCAAGGT-----GGTGATAC 2092
tktA3        ------------------------------------------------------------

tktA1        TCGCTGACG-GCGACCATCGACCCCAGGCACCTCGGCGCACTGATCGCCCTGTACGAGCA 2156
tktA2        TCGATCCGGAGCGGCCGCATA--------------------------------------- 2113
tktA3        ------------------------------------------------------------

tktA1        CTGGATGTTGCGGCCGCTAT 2176
tktA2        -------------------
tktA3        -------------------
```

G.3 A multiple alignemnt of three transketolase genes *tktA1, tktA2* and *tktA3* to *tktA* and *tktB* from *E. coli*. The alignment shows that *tktA* of *E. coli* is unlike those of *A. vinelandii*, but *tktB* appears to have higher similarity score. The *tktA* has a very low similarity with *tktB* of *E. coli*.

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	tktA	1992	2	tktB	2004	1
1	tktA	1992	3	tktA1	2176	2
1	tktA	1992	4	tktA2	2113	2
1	tktA	1992	5	tktA3	1998	1
2	tktB	2004	3	tktA1	2176	64
2	tktB	2004	4	tktA2	2113	64
2	tktB	2004	5	tktA3	1998	66
3	tktA1	2176	4	tktA2	2113	98
3	tktA1	2176	5	tktA3	1998	75
4	tktA2	2113	5	tktA3	1998	75

```
tktA1        GATAACGCATATGACCCAGAACCTGAAACCGGGCCGACGCGCCCTGGCCAATGCCATCCG 60
tktA2        GATAACGCATATGACCCAGAACCTGAAACCGGGCCGACGCGCCCTGGCCAATGCCATCCG 60
tktA3                       ATGCCCAGCCGTCGTGAGCGAGCCAATGCCATTCG 35
tktB                       ATGTCCCGAAAAGACCTTGCCAATGCGATTCG 32
tktA                       TTACAGCAGTTCTTTTGCTTTCGCAA-CAA 29

tktA1        CGTCCTGAGCATGGATGCGGTGGAACGGGCGAACTCCGGCCATCCCGGCGCCCCCATGGG 120
tktA2        CGTCCTGAGCATGGATGCGGTGGAACGGGCGAACTCCGGCCATCCCGGCGCCCCCATGGG 120
tktA3        TGCCCTGAGTATGGATGCCGTGCAGAAAGCCAACAGTGGCCATCCTGGCGCCCCCATGGG 95
tktB         TGCACTCAGTATGGATGCGGTACAAAAAGCCAATTCTGGTCATCCCGGCGCGCCCGATGGG 92
tktA         CGTTATCAACAGTGAAGC--CGAACTCTTCAAACAGCAGCTCTGCCGGAGCAGATTCACC 87

tktA1        CATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGCGCCACAACCCGGGCAACCC 180
tktA2        CATGGCCGATATCGCCGAAGTGCTCTGGAACGGCCACCTGCGCCACAACCCGGGCAACCC 180
tktA3        CATGGCGGATATCGCCGAAGTGCTGTGGCACGACTATCTCAGGCACAACCCGGCCAATCC 155
tktB         CATGGCTGATATTGCCGAAGTGCTGTGGAACGATTTTCTTAAACATAACCCCACCGACCC 152
tktA         GAAGGTGGTCAT-ACCGACG----ATAGCACCGTTCAGGCCAACATACTTGTACCAGTAG 142

tktA1        CCGCTGGC-CCGACCGCGACCGCTTCGTGCTCTCCAACGGGCACGGCTCGATGCTGCTCT 239
tktA2        CCGCTGGC-CCGACCGCGACCGCTTCGTGCTCTCCAACGGGCACGGCTCGATGCTGCTCT 239
tktA3        GCAATGGG-CCGACCGCGACCGCTTCGTGCTCTCCAACGGCCATGGCTCGATGCTGGTCT 214
tktB         AACCTGGT-ATGACCGTGACCGCTTTATTCTTTCCAACGGTCACGCGTCGATGCTGCTCT 211
tktA         TCAGCAATACCCGCTTCTACAGCAACGCGTGCAGTAACCGCTTTCGG-CAGTACGGATTC 201

tktA1        ACAGCCTGCTGCACCTCTCCCGGCTACGACTTGCCGGTCGAGGAGCTCAAGCGCTTCCGCC 299
tktA2        ACAGCCTGCTGCACCTCTCCCGGCTACGACTTGCCGGTCGAGGAGCTCAAGCGCTTCCGCC 299
tktA3        ATTCGCTGCTGCATCTGACCGGCTACGACCTCTCGATCGACGACCTGAAGAACTTCCGCC 274
tktB         ACAGTTTGCTGCATCTGACCGGTTACGACCTGCCACTGGAAGAACTGAAGAACTTCCGTC 271
tktA         ACGGTAAGCAGCATCCTGCTTGTCA-AATGCGTCGGTAGACGGCATGGA-------CACC 253
```

```
tktA1    AGTTGCATTCGCGCACTCCCGGGCACCCGGAGTACGGCTACACGCCGGGGGTGGAGACCA 359
tktA2    AGTTGCATTCGCGCACTCCCGGGCACCCGGAGTACGGCTACACGCCGGGGGTGGAGACCA 359
tktA3    AGTTCGGCAGCCGCACCCCCGGCCATCCCGAATACGGCTATACCCCCGGCGTCGAGACCA 334
tktB     AGTTGCATTCGAAAACCCCTGGTCACCCGGAAATCGGCTATACGCCAGGTGTTGAAACCA 331
tktA     ACGCGCGCTTTCACGCCTTCGGCAGTCAGTTTTTCG--TAGGCAGCAACAGCCAGTTCAA 311

tktA1    CCACCGGTCCGCT--CGGCCAGGGGCTGGCCAACGCCGTCGGCATGG-CCCTGGCCGAGC 416
tktA2    CCACCGGTCCGCT--CGGCCAGGGGCTGGCCAACGCCGTCGGCATGG-CCCTGGCCGAGC 416
tktA3    CCACCGGCCCGCT--CGGCCAGGGCCTGGCCAACGCCGTCGGCATGG-CCGTGGCCGAGA 391
tktB     CCACCGGTCCTCT--TGGACAAGGTTTGGCGAACGCCGTCGGGCTGG-CGATAGCAGAGC 388
tktA     CTTCTGAACCGGTAGCGATGAAAATCAGTTCCGGCTGACCGGCGCAGTCTTTCAGCACAT 371

tktA1    GGCTGCTGGCGGCGGACTTCAACCGCCCGGGACAGGCGATCGTCGACCACCGCACCTACG 476
tktA2    GGCTGCTGGCGGCGGACTTCAACCGCCCGGGACAGGCGATCGTCGACCACCGCACCTACG 476
tktA3    AGGTGCTGGCCGCGCAGTTCAACCGGCCCGGGCAGCGCATCGTCGATCATTACACCTACG 451
tktB     GTACGCTGGCGGCGCAATTTAATCAGCCGGATCATGAAATTGTCGATCACTTCACCTATG 448
tktA     AACCACCGCGCGCGATGTTTGCCAGTT--GCTCTTCAGTTCGTTC-CTGCTGCGCC-AGG 427

tktA1    CCTTCGTCGGCGACGGCT-GCCTGATGGAGGGCATCTCCCACGAGGTCTGCTCGCTGGCC 535
tktA2    CCTTCGTCGGCGACGGCT-GCCTGATGGAGGGCATCTCCCACGAGGTCTGCTCGCTGGCC 535
tktA3    TGTTCCTCGGCGACGGCT-GCCTGATGGAAGGCATCTCCCACGAAGTCTGCTCGCTGGCC 510
tktB     TGTTTATGGGCGACGGCT-GCCTGATGGAAGGTATCTCTCACGAAGTCTGTTCGCTGGCC 507
tktA     TTCTGACGGGAGAGGATCAGTGCGGTCGGGCCGTCCTGACGCTCAACACCGTATTTCCAC 487

tktA1    GGCACCCTGGGACTCGGCAAGCTGATCGTGTTCTACGACGACAACGGCATCTCCATCGAC 595
tktA2    GGCACCCTGGGACTCGGCAAGCTGATCGTGTTCTACGACGACAACGGCATCTCCATCGAC 595
tktA3    GGCACCCTGGGGCTGAACAAGCTGGTCGCCTTCTACGACGACAACGGCATCTCCATCGAC 570
tktB     GGCACGCTGGGGCTGGGCAAGCTGATTGGTTTTTACGATCACAACGGTATTTCCATCGAC 567
tktA     GCGACCGCGGATTCAACCTGGTCACACGGACGCCATGTAGACATGTTCGGGGTTACGCGC 547

tktA1    GGTGAGGTCGCGGGCTGGTTCAGCGAC-GACACGCCGGCGCGCTTCCGCGCCTACGGCTG 654
tktA2    GGTGAGGTCGCGGGCTGGTTCAGCGAC-GACACGCCGGCGCGCTTCCGCGCCTACGGCTG 654
tktA3    GGCGAGGTCGAGGGCTGGTTCACCGAC-GACACCCCGGCGCGCTTCGAGGCCTACGGCTG 629
tktB     GGTGAAACCGAAGGCTGGTTTACCGAC-GATACGGCAAAACGTTTTGAAGCCTATCACTG 626
tktA     AGAGAAGC---GACCTGCTCTACCGGCTGGTGAGTCGGGCCATCTTCGCCCAGACCGATG 604

tktA1    GCAGGTCGTCGAGGCGGTCGACGGCCACGACGCCAGCGCGCTCGAGGCGGCGCTGGCGGC 714
tktA2    GCAGGTCGTCGAGGCGGTCGACGGCCACGACGCCAGCGCGCTCGAGGCGGCGCTGGCGGC 714
tktA3    GCTGGTGATCCGCAACGTCGACGGCCACGACGCCGACGAGATCCGGACCGCGATCGAGAC 689
tktB     GCATGTGATCCATGAAATCGACGGCCACGATCCGCAGGCGGTGAAGGAAGCGATCCTTGA 686
tktA     G--AGTCGTGGGTGTAAACCATCACC-TGACGCTGTTTCA-TCAGCGCAGCCATACGTAC 660

tktA1    GGCCAAGGCCG-ATACGGAGCGGCCGACGCTGATCTGCTGCAAGACCACCATCGGCTTCG 773
tktA2    GGCCAAGGCCG-ATACGGAGCGGCCGACGCTGATCTGCTGCAAGACCACCATCGGCTTCG 773
tktA3    CGC---GCGCA-AGAGCGAGCGCCCGGTCCTGATCTGCTGCAAGACGGTGATCGGTTTCG 745
tktB     AGCGCAAAGCG-TGAAAGATAAGCCGTCGCTGATTATCTGCCGTACGGTGATTGGCTTTG 745
tktA     GGCGTTACGTGCGTATTCCACGAACATCAGGAAGGTGGAGGTGTACGGCAGGAAACCACC 720

tktA1    GCGCGCCGACCAAGGCCGGCAGCCACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGA 833
tktA2    GCGCGCCGACCAAGGCCGGCAGCCACGACTGCCACGGCGCCCCGCTGGGCGGCGAGGAGA 833
tktA3    GTTCGCCGAACAAGCAGGGCAAGGAAGCCTGCCACGGCGCCGCGCTGGGGGCGGAGGAAA 805
tktB     GTTCGCCGAATAAAGCAGGTAAGGAAGAGGCGCACGGCGCACCGCTGGGGGAAGAAGAAG 805
tktA     GTGCAGGGA--GATACCGTTAGCAATCGCGGTCATACCGAACTCGCGAACACCGTAGTGG 778

tktA1    TCGCCGCGGCCCGCGAGGCGCTGGACTGGCCGCACGCGCCCTTCGAGGTGCCCGCGCAGA 893
tktA2    TCGCCGCGGCCCGCGAGGCGCTGGACTGGCCGCACGCGCCCTTCGAGGTGCCCGCGCAGA 893
tktA3    TCGCCCTGACCCGCGACGCGCTGGGCTGGAAGCACGCGCCGTTCGAGATTCCCGCCGAGA 865
tktB     TAGCTCTGGCACGGCAAAAACTGGGCTGGCACCATCCGCCATTTGAGATCCCTAAAGAGA 865
tktA     ATGTAGTTACCCGCAGCAT-CTTCGTTGATTGCTTTAG--AACCAGACCACAGGGTCAGG 835

tktA1    TCGCCAGGGCCTGGGACGCGCGCGAGCGCGGGGCGGCGCTGGAGCGGGACTGGCGACAAC 953
tktA2    TCGCCAGGGCCTGGGACGCGCGCGAGCGCGGGGCGGCGCTGGAGCGGGACTGGCGACAAC 953
tktA3    TCTACGCCGGCTGGGACGCCAGGGCCGCCGGCGCCGCCCGCGAGACCGAATGGAACCGGC 925
tktB     TTTATCACGCCTGGGATGCCCGTGAAAAAGGCGAAAAAGCGCAGCAGAGCTGGAATGAGA 925
tktA     TTAGACGGTGCCAG---GTCAGCAGAACCGCCGAGGAATTCCGGCAACAGCGGACCGAAA 892

tktA1    GCTTCGACGAC--TACGCGCGGGCCTAC--CCGGCGCTGGCCGCCGAACTGGAGCGGCGC 1009
tktA2    GCTTCGACGAC--TACGCGCGGGCCTAC--CCGGCGCTGGCCGCCGAGCTGGAGCGGCGC 1009
tktA3    GCTTCGCCGCC--TACCAGGCCGAGTTC--CCGGAACTGGCCGCGGGAGTTCCTGCGCCGC 981
tktB     AGTTTGCCGCC--TATAAAAAGGCTCAT--CCGCAACTGGCAGAAGAGTTTACCCGTCGG 981
tktA     GCTTCGATAGCATTCTGAGACGCTTTACGGCTGGCGATTTTCGCCGGGTTAGCCTG-CAG 951

tktA1    CTGGCCGGCGAGCTGCCGGCCGACTGGGGCGCGCGGCGCGACGCGCTGATCGCCGG--CT 1067
tktA2    CTGGCCGGCGAGCTGCCGGCCGACTGGGGCGCGCGGCGCGACGCGCTGATCGCCGG--CT 1067
tktA3    AGCCGTGGCGAACTGCCCGCCGATTTCGCCGCGAAGGCCGGCGAATACATCGCCGAAATC 1041
tktB     ATGAGCGGCGGTTTGCCGAAGGACTGGGAGAAAAACGACTCAGAAATATATCAATGAG-TT 1040
tktA     TTTAGCAATGAACT--CTTTTGCTTTCGCGTCGAAGTCAGACG-GCATTTCGCCTTTCAT 1008
```

112

```
tktA1    GCCAGCAGGCGGCGCGCGAGCTGGCGACCCGCAAGAGCTCGCAGCAGGTCCTGGAAGCGC 1127
tktA2    GCCAGCAGGCGGCGCGCGAGCTGGCGACCCGCAAGAGCTCGCAGCAGGTCCTGGAAGCGC 1127
tktA3    GCCGCCAAGGGGGAAACCATC--GCCAGCCGCAAGGCCAGCCAGAACGCGCTGAACGCCT 1099
tktB     GCAGGCGAATCCGGCGAAAATC-GCTACCCGTAAGGCTTCGCAAAATACGCTTAACGCTT 1099
tktA     ACGGCGGGTAAATTCAGCGGCTTCCTGCGGATAAGCTTTCGCGTAAGCAGCGAATTTCTC 1068

tktA1    TGGGCCCGCTGCTGCCGGAGC-TGCTCGGCGGCTCGGCGGACCTGGCGCCCTCGAACCTG 1186
tktA2    TGGGCCCGCTGCTGCCGGAAC-TGCTCGGCGGCTCGGCGGACCTGGCGCCCTCGAACCTG 1186
tktA3    TCGGCCCGCTGCTGCCGGAGT-TCCTCGGCGGCTCGGCGGACCTGGCCGGTTCCAACCTG 1158
tktB     ACGGGCCGATGCTGCCTGAGT-TGCTCGGCGGTTCGGCGGATCTGGCTCCCAGCAACCTG 1158
tktA     GTTCCATGCGGATTCTTTCGCCTGGCCTGCTTCTTTCGCATCCCACTGAGCATAGATTTC 1128

tktA1    ACCCGCTGGTCCGGCTCCCGCTCGGTCGGCGGCGAGGCGCCCGGGGGCAACTACATCCAC 1246
tktA2    ACCCGCTGGTCCGGCTCCCGCTCGGTCGGCGGCGAGGCGCCCGGGGGCAACTACATCCAC 1246
tktA3    ACCCTGTGGAAGGGCTGCAAGGCGCTCGCGGCGGACGACGCCTCCGGTAACTACCTGCAC 1218
tktB     ACCATCTGGAAAGGTTCTGTTTCGCT---GAAGGAAGATCCGGCGGGCAACTACATTCAC 1215
tktA     AGACGGGATTTCGAACGGTGCATATTTCCAGCCCAGCTGTTCGCGGGTCAGGGCAATTTC 1188

tktA1    TACGGGGTGCGCGAGTTCGGCATGAGCGCGATGATGAACGGGCTGGCGCTGCACGGCGGT 1306
tktA2    TACGGGGTGCGCGAGTTCGGCATGAGCGCGATGATGAACGGGCTGGCGCTGCACGGCGGT 1306
tktA3    TACGGCGTGCGCGAGTTCGGCATGAGCGCGATCATGAACGGTATCGCTCTGCACGGCGGC 1278
tktB     TACGGGGTGCGTGAATTTGGTATGACCGCTATCGCCAACGGCATCGCGCACCACGGCGGC 1275
tktA     AGCGTCGCCCAGCGGCGCACCGTGGGAGTCGTGAGTACCGGCTTTGTTCGGGGAACCGAA 1248

tktA1    TTCATCCCCTACGGCGGCACCTTCCTGATGTTCATGGAATACGC-CCGCA-ACGCCGTGC 1364
tktA2    TTCATTCCCTACGGCGGCACCTTCCTGATGTTCATGGAATACGC-CCGCA-ACGCCGTGC 1364
tktA3    TTCATTCCCTACGGCGCCACCTTCCTGGTGTTCATGGAGTACGC-CTGCA-ACGCCGTGC 1336
tktB     TTTGTGCCGTATACCGCGACGTTCCTGATGTTTGTTGAATACGC-CCGTA-ACGCCGCGC 1333
tktA     ACCGATGATGGTTTTG-CACATCAGCAGGGACGGTTTGTCAGTCACTGCGCGCGCTTCTT 1307

tktA1    GCATGGCCGCGCTGATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCC 1424
tktA2    GCATGGCCGCGCTGATGAAGCTGCGCGCGGTGTTCGTCTACACCCACGACTCGATCGGCC 1424
tktA3    GCATGGCCGCGCTGATGAAGCAGCGCGTGCTGTTCGTCTACACCCACGACTCCATCGGCC 1396
tktB     GGATGGCGGCACTGATGAAAGCGCGGCAGATTATGGTTTATACCCACGACTCAATTGGTC 1393
tktA     CTACTGCGCGTTTGATGGATGC-CGCGTCATGACCGTCGATGTC-GCGAAT--AACGTGC 1363

tktA1    TGGGCGAGGACGGCCCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCG-CCAGACG 1483
tktA2    TGGGCGAGGACGGCCCGACCCACCAGCCGGTCGAACAGCTCGCCAGCCTGCG-CCAGACG 1483
tktA3    TCGGCGAGGACGGCCCGACCCACCAGCCGGTCGAGCAACTGGCCAGCCTGCG-CAGCACG 1455
tktB     TGGGCGAAGATGGTCCGACGCACCAGGCTGTTGAGCAACTGGCTAGTCTGCG-CTTAACG 1452
tktA     CAGCCGTAAGCTTCGAAACGCATTGCGGTGTCGT-CGGTGAACCAGCCTTCAACGTGACC 1422

tktA1    CCGAACCTGGAAACC-TGGCGGCCCTGCGACGAGACGG----AGACCGCGGTGGCCTGGT 1538
tktA2    CCGAACCTGGAAACC-TGGCGGCCCTGCGACGAGACGG----AGACCGCGGTGGCCTGGT 1538
tktA3    CCGAACCTCGACACC-TGGCGCCCCGTGCGACGCGGTGG----AGTCGGCGGTGGCCTGGA 1510
tktB     CCAAATTTCAGCACC-TGGCGACCGTGCGATCAGGTGG----AAGCGGCGGTTGGCTGGA 1507
tktA     ATCAATAGAGATACCGTTGTCATCGTAGAACGCAATCAGTTTACCCAGCTTCAGCGTACC 1482

tktA1    GCGCGGCGCTCGAACGGCG--GGACGGCCCGTCGGCGCTGGTGCTGTCGCGGCAGAA--C 1594
tktA2    GCGCGGCGCTCGAACGGCG--GGACGGCCCGTCGGCGCTGGTGCTGTCGCGGCAGAA--C 1594
tktA3    AATGCGCCATCGAGCGCCA--GGACGCCCGGCGACCACCCTGGTGTTCAGCCGCCAGAA--C 1566
tktB     AGCTGGCGGTTGAGCGCCA--CAACGGACCGACAGCCACTGATCCTCTCAAGGCAGAA--T 1563
tktA     CGCCAGAGAGCAAACTTCGTGGGAGATGCCTTCCATCATGCAGCCGTCGCCCATGAAGGC 1542

tktA1    CTGCCGGCCGCGGCGCGCGACGCCGGG-CAGCGCGCGGCCATCGAGCGCGGCGCCTATGT 1653
tktA2    CTGCCGGCCGCGGCGCGCGACGCCGGG-CAGCGCGCGGCCATCGAGCGCGGCGCCTATGT 1653
tktA3    CTGCCGCACCAGGCGCGCGATGCCCGG-CAACTGGCCGATATCGCCCGCGGCGGCTACGT 1625
tktB     CTGGCACAGGTAGAACG-AACACCGGATCAGGTTAAAGAGATTGCTCGTGGCGGTTATGT 1622
tktA     ATAGGTGTAGTGGTCGACAATGTCGTGGCC-CGGACGGTTAAACTGCGCCGCCAGCGTTT 1601

tktA1    GCTGCTCGACTGCCCGGGC-ACGCCGGAGCTGATCCTGCTGGCGTCCGGCTCCGAGGTGG 1712
tktA2    GCTGCTCGACTGCCCGGGC-ACGCCGGAGCTGATCCTGCTGGCGTCCGGCTCCGAGGTGG 1712
tktA3    GCTCCGCGATTGTGCCGGC-ACCCCGGAGCTGATCCTCATCGCCACCGGTTCGGAGGTCG 1684
tktB     GCTGAAAGACAGCGGCGGT-AAGCCAGATATTATTCTGATTGCCACCGGTTCAGAGATGG 1681
tktA     TTTCTGCAATCGCCATACCGACTGCGTTGGCAATACCCTGACCCAGCGGACCGGTGGTGG 1661

tktA1    CGCTGGCGCTGGCGGCGGCGCAGGCG-CTGGGCGAGGCGG-GCCGGGCGGTCCGCGTGGT 1770
tktA2    CGCTGGCGCTGGCGGCGGCGCAGGCG-CTGGGCGAGGCGG-GCCGGGCGGTCCGCGTGGT 1770
tktA3    CCCTGGCGGTGCAGGCCCACGAGACG-CTGAGCGGGCAGG-GCCGCCGGGTGCGAGTGGT 1742
tktB     AAATCACCCTGCAAGCGGCAGAGAAA-TTAGCGGGAGAAG-GTCGCAATGTTCGCGTAGT 1739
tktA     TTTCCACACCAGCGGTGTAACCCACTTCCGGGTGACCCGGAGTTTTAGAGTGCAGCTGAC 1721
```

113

```
tktA1    GTCGGTGCCGT-GCCTGGAGCGTTTCGAGCGGCAGGCGGCGGACTACCGCGAGGCGGTGC 1829
tktA2    GTCGGTGCCGT-GCCTGGAGCGTTTCGAGCGGCAGGCGGCGGACTACCGCGAGGCGGTGC 1829
tktA3    TTCCATGCCCTCGACCAGCGTGTT-CGATCGCCAGGACGCCGCCTGGAAACAGGCCGTGC 1801
tktB     TTCCCTGCCCTCGACCGATATTTT-CGATGCCCAGGATGAGAAATATCGGGAGTCGGTAC 1798
tktA     GGAAGTTTTTCAGTTCTTCCATCGGCAGATCGTAACCGGTGAGGTGCAGC-AGGCTGTAG 1780

tktA1    TGCCGTCCGCGGTGCGGGCCCGCGTCGCGGTCGAGATGGCGCGTCCGGAGAGCTGGCACC 1889
tktA2    TGCCGTCCGCGGTGCGGGCCCGCGTCGCGGTCGAGATGGCGCGTCCGGAGAGCTGGCACC 1889
tktA3    TGCCGTCCGGGGTCGGGGCGCGGATCGCCATCGAGGCCGCCCACGCCGACTTCTGGTACA 1861
tktB     TGCCTTCAAATGTTTCTGCCCGCGTGGCAGTAGAAGCGGGTATTGCCGATTACTGGTACA 1858
tktA     ATCAGCATGGAGCCGTGGC-CGTTGGACAGCACGAAGCGGTCACGGTCAGCCCAGGACGG 1839

tktA1    GCTACGTGGGGCTCGACGGCGCGGTGCTCGGCATGGAGGGGTTCGGCGA-GTCGGCGCCG 1948
tktA2    GCTACGTGGGTCTCGACGGCGCGGTGCTCGGCATGGAGGGGTTCGGCGA-GTCGGCGCCG 1948
tktA3    AGTACGTCGGGCTGGCTGGACGCATCGTCGGCATGACCCGCTTCGGCGA-GTCCGCGCCG 1920
tktB     AATATGTTGGCCTGAAAGGGGCAATTGTCGGGATGACAGGTTATGGGGA-ATCAGCTCCG 1917
tktA     ATTCTGCGGGTTGTGTTTCAGGAAATCACGCCACAG--GACTTCGGCAATGTCAGCCATA 1897

tktA1    GCGGCGGAGCTGTTCCGGCATTTCGGCTTCAGCGTCGAGAACCTGC---TCGCCAGGGCG 2005
tktA2    GCGGCGGAGCTGTTCCGGCATTTCGGCTTCAGCGTCGAGAACCTGC---TCGCCAGGGCG 2005
tktA3    GCCGGCGCGCTCTTCGAGCATTTCGGCTTCACCCTCGACAACCTGC---TGGCCAGCGCC 1977
tktB     GCGGATAAGCTGTTCCCGTTCTTTGGCTTTACCGCCGAGAATATTG---TGGCAAAAGCG 1974
tktA     CCCATAGGGGCACCCGGGTGACCGGATTTGGCTTTCTGTACTGCGTCCATGCTCAGCGCA 1957

tktA1    CAGGCCGTGCTGGCC-GGGCTGGAGGGGAGGGCGTAGCCATGTATCTGGGCGTCGATTGC 2064
tktA2    CAGGCCGTGCTGGCC-GGGCTGGAGGGGAGGGCGTAGCCATGTATCTGGGCGTCGATTGC 2064
tktA3    GGGGAACTGCTCGG--GGACTGA-------------------------------------1998
tktB     CATAAGGTACTGGGA-GTGAAAGGTGCCTGA----------------------------- 2004
tktA     CGAATAGCATTGGCAAGCTCTTTACGTGAGGACAT------------------------- 1992

tktA1    GGGACCCAGGGGACCAAACCAGCCGGGTAACACTCGCTGACG-GCGACCATCGACCCCAG 2123
tktA2    GGGACCCAGGGGACCAAGGT-----GGTGATACTCGATCCGGAGCGGCCGCATA------ 2113
tktA3    ------------------------------------------------------------
tktB     ------------------------------------------------------------
tktA     ------------------------------------------------------------


tktA1    GCACCTCGGCGCACTGATCGCCCTGTACGAGCACTGGATGTTGCGGCCGCTAT 2176
tktA2    ----------------------------------------------------
tktA3    ----------------------------------------------------
tktB     ----------------------------------------------------
tktA     ----------------------------------------------------
```

G.4 A pair wise alignemnt of three transketolase genes *tktA1, tktA2* and *tktA3* to thier deletion mutants.

Transketolase 1, the wild-type and the mutant pair wise aligned, the deletion occur from 189 to 672 aa.

```
mutant    LKPGRRALANAIRVLSMDAVERANSGHPGAPMGMADIAEVLWNGHLRHNPGNPRWPDRDR 60
WT        LKPGRRALANAIRVLSMDAVERANSGHPGAPMGMADIAEVLWNGHLRHNPGNPRWPDRDR 60
          ************************************************************

mutant    FVLSNGHGSMLLYSLLHLSGYDLPVEELKRFRQLHSRTPGHPEYGYTPGVETTTGPLGQG 120
WT        FVLSNGHGSMLLYSLLHLSGYDLPVEELKRFRQLHSRTPGHPEYGYTPGVETTTGPLGQG 120
          ************************************************************

mutant    LANAVGMALAERLLAADFNRPGQAIVDHRTYAFVGDGCLMEGISHEVCSLAGTLGLGKLI 180
WT        LANAVGMALAERLLAADFNRPGQAIVDHRTYAFVGDGCLMEGISHEVCSLAGTLGLGKLI 180
          ************************************************************

mutant    VFYDDNGIS-------------------------------------------------- 189
WT        VFYDDNGISIDGEVAGWFSDDTPARFRAYGWQVVEAVDGHDASALEAALAAAKADTERPT 240
          *********

mutant    ------------------------------------------------------------
WT        LICCKTTIGFGAPTKAGSHDCHGAPLGGEEIAAAREALDWPHAPFEVPAQIARAWDARER 300
```

```
mutant        ------------------------------------------------------------
WT            GAALERDWRQRFDDYARAYPALAAELERRLAGELPADWGARRDALIAGCQQAARELATRK 360


mutant        ------------------------------------------------------------
WT            SSQQVLEALGPLLPELLGGSADLAPSNLTRWSGSRSVGGEAPGGNYIHYGVREFGMSAMM 420


mutant        ------------------------------------------------------------
WT            NGLALHGGFIPYGGTFLMFMEYARNAVRMAALMKLRAVFVYTHDSIGLGEDGPTHQPVEQ 480


mutant        ------------------------------------------------------------
WT            LASLRQTPNLETWRPCDETETAVAWCAALERRDGPSALVLSRQNLPAAARDAGQRAAIER 540


mutant        ------------------------------------------------------------
WT            GAYVLLDCPGTPELILLASGSEVALALAAAQALGEAGRAVRVVSVPCLERFERQAADYRE 600


mutant        ------------------------------------------------------------
WT            AVLPSAVRARVAVEMARPESWHRYVGLDGAVLGMEGFGESAPAAELFRHFGFSVENLLAR 660


mutant        ------------
WT            AQAVLAGLEGRA 672
```

Transketolase 2, the wild-type and the mutant pair wise aligned, the deletion occur from 189 to 672 aa.

```
mutant        LKPGRRALANAIRVLSMDAVERANSGHPGAPMGMADIAEVLWNGHLRHNPGNPRWPDRDR 60
WT            LKPGRRALANAIRVLSMDAVERANSGHPGAPMGMADIAEVLWNGHLRHNPGNPRWPDRDR 60
              ************************************************************

mutant        FVLSNGHGSMLLYSLLHLSGYDLPVEELKRFRQLHSRTPGHPEYGYTPGVETTTGPLGQG 120
WT            FVLSNGHGSMLLYSLLHLSGYDLPVEELKRFRQLHSRTPGHPEYGYTPGVETTTGPLGQG 120
              ************************************************************

mutant        LANAVGMALAERLLAADFNRPGQAIVDHRTYAFVGDGCLMEGISHEVCSLAGTLGLGKLI 180
WT            LANAVGMALAERLLAADFNRPGQAIVDHRTYAFVGDGCLMEGISHEVCSLAGTLGLGKLI 180
              ************************************************************

mutant        VFYDDNGISIDGEVAGWFSDDTPARFRAYGWQVVEAVDGHDASALEAALAAAKADTERPT 240
WT            VFYDDNGISIDGEVAGWFSDDTPARFRAYGWQVVEAVDGHDASALEAALAAAKADTERPT 240
              ************************************************************

mutant        LICCKTTIGF-------------------------------------------------- 250
WT            LICCKTTIGFGAPTKAGSHDCHGAPLGGEEIAAAREALDWPHAPFEVPAQIARAWDARER 300
              **********

mutant        ------------------------------------------------------------
WT            GAALERDWRQRFDDYARAYPALAAELERRLAGELPADWGARRDALIAGCQQAARELATRK 360


mutant        ------------------------------------------------------------
WT            SSQQVLEALGPLLPELLGGSADLAPSNLTRWSGSRSVGGEAPGGNYIHYGVREFGMSAMM 420


mutant        ------------------------------------------------------------
WT            NGLALHGGFIPYGGTFLMFMEYARNAVRMAALMKLRAVFVYTHDSIGLGEDGPTHQPVEQ 480


mutant        ------------------------------------------------------------
WT            LASLRQTPNLETWRPCDETETAVAWCAALERRDGPSALVLSRQNLPAAARDAGQRAAIER 540


mutant        ------------------------------------------------------------
WT            GAYVLLDCPGTPELILLASGSEVALALAAAQALGEAGRAVRVVSVPCLERFERQAADYRE 600


mutant        ------------------------------------------------------------
WT            AVLPSAVRARVAVEMARPESWHRYVGLDGAVLGMEGFGESAPAAELFRHFGFSVENLLAR 660
```

```
mutant           ------------
WT               AQAVLAGLEGRA 672
```

Transketolase 3, the wild-type and the mutant pair wise aligned, the deletion occur from 1 to 440 aa.

```
mutant           ------------------------------------------------------------
WT               PSRRERANAIRALSMDAVQKANSGHPGAPMGMADIAEVLWHDYLRHNPANPQWADRDRFV 60


mutant           ------------------------------------------------------------
WT               LSNGHGSMLVYSLLHLTGYDLSIDDLKNFRQFGSRTPGHPEYGYTPGVETTTGPLGQGLA 120


mutant           ------------------------------------------------------------
WT               NAVGMAVAEKVLAAQFNRPGQRIVDHYTYVFLGDGCLMEGISHEVCSLAGTLGLNKLVAF 180


mutant           ------------------------------------------------------------
WT               YDDNGISIDGEVEGWFTDDTPARFEAYGWLVIRNVDGHDADEIRTAIETARKSERPVLIC 240


mutant           ------------------------------------------------------------
WT               CKTVIGFGSPNKQGKEACHGAALGAEEIALTRDALGWKHAPFEIPAEIYAGWDARAAGAA 300


mutant           ------------------------------------------------------------
WT               RETEWNRRFAAYQAEFPELAAEFLRRSRGELPADFAAKAGEYIAEIAAKGETIASRKASQ 360


mutant           ------------------------------------------------------------
WT               NALNAFGPLLPEFLGGSADLAGSNLTLWKGCKALAADDASGNYLHYGVREFGMSAIMNGI 420


mutant           ----------ATFLVFMEYACNAVRMAALMKQRVLFVYTHDSIGLGEDGPTHQPVEQLAS 50
WT               ALHGGFIPYGATFLVFMEYACNAVRMAALMKQRVLFVYTHDSIGLGEDGPTHQPVEQLAS 480
                           **************************************************

mutant           LRSTPNLDTWRPCDAVESAVAWKCAIERQDGPSTLVFSRQNLPHQARDARQLADIARGGY 110
WT               LRSTPNLDTWRPCDAVESAVAWKCAIERQDGPSTLVFSRQNLPHQARDARQLADIARGGY 540
                 ************************************************************

mutant           VLRDCAGTPELILIATGSEVALAVQAHETLSGQGRRVRVVSMPSTSVFDRQDAAWKQAVL 170
WT               VLRDCAGTPELILIATGSEVALAVQAHETLSGQGRRVRVVSMPSTSVFDRQDAAWKQAVL 600
                 ************************************************************

mutant           PSGVGARIAIEAAHADFWYKYVGLAGRIVGMTRFGESAPAGALFEHFGFTLDNLLASAGE 230
WT               PSGVGARIAIEAAHADFWYKYVGLAGRIVGMTRFGESAPAGALFEHFGFTLDNLLASAGE 660
                 ************************************************************

mutant           LLGD 234
WT               LLGD 664
                 ****
```

116

Appendix H Alignment of PCR products in designated plasmids

The pair wise alignment carried out on Clone Manager of the inserted PCR fragments in the designated plasmid vectors to the transketolase genes obtained from the *A. vinelandii* genome and the genes designated in Clone Manager. The primers used in the sequencing reaction are shown in table 2.4.

H.1 Alginemnt of PCR-*tktA1* inserted into pMK2 to the *tktA1* genes from *A. vinelandii* genome sequence.

H.2 Alginemnt of PCR-*tktA2* inserted into pMK6 to the *tktA2* genes from *A. vinelandii* genome sequence

H.3 Alginemnt of PCR-*opptkA3* inserted into pMK11 to the *opptktA3* sequence designed in Clone Manager.

H.4 Alginemnt of PCR-*opptktA2* inserted into pHE304 to the *opptktA2* sequence designed in Clone Manager.

H.1 Alignment of *tktA1* insert and the *tktA1* gene from *A. vinelandii*.

tktA12ATG   759
Pm_forw   1070
tktA24767   198

Insert_tktA1   943
tktA1oppR   26
M13rev   1190
tktA1NotIR   1154
tktA12ATG   909
Pm_forw   1095
tktA24767   348

Insert_tktA1 1089
tktA1oppR   26
M13rev   1185
tktA1NotIR   1054
tktA12ATG   1026
Pm_forw   1095
tktA24767   494

Insert_tktA1 1238
tktA1oppR   26
M13rev   1043
tktA1NotIR   907
tktA12ATG   1061
Pm_forw   1193
tktA24767   643

Insert_tktA1 1388
tktA1oppR   26
M13rev   896
tktA1NotIR   757
tktA12ATG   1068
Pm_forw   1193
tktA24767   793

Insert_tktA1 1538
tktA1oppR   26
M13rev   746
tktA1NotIR   607
tktA12ATG   1128
Pm_forw   1193
tktA24767   943

Insert_tktA1 1687
tktA1oppR   0
M13rev   597
tktA1NotIR   458
tktA12ATG   1128
Pm_forw   1193
tktA247671   091

Insert_tktA1 1837
tktA1oppR   0
M13rev   447
tktA1NotIR   308
tktA12ATG   1128
Pm_forw   1193
tktA24767   1216

Insert tktA1 1987 gaacctgctcgccaggcggcgcaggccgtgctggccggctggaggggaggggcgtagccatgtatctggcgtcgtcgattgcgtcgtcgacacactcgctgacggcgcgaccatcgacccagccagcacctcggcgcca
tktAloppR   0 ---
M13rev   297 gaacctgctcgccaggcggcgcaggccgtgctggccggctggaggggaggggcgtagccatgtatctggcgtcgtcgattgcgtcgtcgacacactcgctgacggcgcgaccatcgacccagccagcacctcggcgcgca
tktA1NotIR   158 gaacctgctcgccaggcggcgcaggccgtgctggccggctggaggggaggggcgtagccatgtatctggcgtcgtcgattgcgtcgtcgacacactcgctgacggcgcgaccatcgacccagccagcacctcggcgcca------
tktA12ATG   1128 --acctgctcgccaggcggcgcaggccgtgctggccggctggaggggaggggcgtagccatgt--
Pm forw   1247 ---
tktA24767   1216 ---

Insert tktA1 2137 ctgatcgccctgtacgagcactggatgttgcggccgctat--
tktAloppR   0 ---
M13rev   147 ctgatcgccctgtacgagcactggatgttgcggccgcacttgtataagagtcagaattcctcgagacgtacggtgcgcgcgatgcattcgaaga--
tktA1NotIR   20 ctgatcgccctgtacgagca---
tktA12ATG   1187 ---
Pm forw   1247 ---
tktA24767   1216 ---

H.2 Alignment of *tktA2* insert and the *tktA2* gene from *A. vinelandii*.

```
Insert tktA2     10  ------------------------------------------------------------------------------------------------------------
tktA2NotIB     1210  ------------------------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------------------------
tktAloppR       737  ------------------------------------------------------------------------------------------------------------
Pm forw         142  atgcccatgcagccagcataaccagcagcataaacgtgtccggtttgataggataagtccagccttgcaagaagcggatacaggagtgcaaaaatgctatcctagaaa
tktA24767         1  ------------------------------------------------------------------------------------------------------------
tktA12ATG         1  ------------------------------------------------------------------------------------------------------------

Insert tktA2     10  ------------------------------tatgaccagaacctgaaacgggccgacgcgcccctgccaatgccatccgcgtcc
tktA2NotIB     1210  ------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------
tktAloppR       737  --tatgaccagaacctgaaacgggccgacgcgcccctgccaatgccatccgcgtcc
Pm forw         252  gacctaccctgaggctctatgcaacatgtacaataataatggagtcatgaacatatgacccagaacctgaaacctgaaacgggccgacgcgcccctgccaatgccatccgcgtcc
tktA24767         1  ------------------------------------------------------------------------------------------
tktA12ATG         1  ----------------------------------cgacgcgcccctgccaatgcc-tccgcgtcc

Insert tktA2     66  tgagcatggatgcggtggaacgggcgaactccggccatccggcgcgccccatgggcatgcggccgatatcgccgatatggccgaagtgctctgaacggccacctgccgcacaacccgggc
tktA2NotIB     1210  ------------------------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------------------------
tktAloppR       681  tgagcatggatgcggtggaacgggcggaactccggccatccggcgcgccccatgggcatgcggccgatatcgccgatatggccgaagtgctctgaacggccacctgccgcacaacccgggc
Pm forw         362  tgagcatggatgcggtggaacgggcggaactccggccatccggcgcgccccatgggcatgcggccgatatcgccgatatggccgaagtgctctgaacggccacctgccgcacaacccgggc
tktA24767         1  ------------------------------------------------------------------------------------------------------------
tktA12ATG        31  tgagcatggatgcggtggaacgggcgaactccggccatccggcgcgccccatgggcatgcggccgatatcgccgatatggccgaagtgctctgaacggccacctgccgcacaacccgggc

Insert tktA2    176  aaccccgctggccgaccgcgaccgcttcgtgctctccaacggcacggctcgatgctcgatctctccggctacgacttgcggtcgaggagct
tktA2NotIB     1210  ------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------
tktAloppR       571  aaccccgctggccgaccgcgaccgcttcgtgctctccaacggcacggctcgatgctgtctgtctcgtctccggctacgactctgcggtcgaggagct
Pm forw         472  aaccccgctggccgaccgcgaccgcttcgtgctctccaacggcacggctcgatgctgtctgtctcgtctccggctacgactctgcggtcgaggagct
tktA24767         1  ------------------------------------------------------------------------------------------
tktA12ATG       141  aaccccgctggccgaccgcgaccgcttcgtgctctccaacggcacggctcgatgctgtctgtctcgtctccggctacgactctgcggtcgaggagct

Insert tktA2    286  caagcgcttccgcgcagttgcattcgcgcactccgggcaccggagtacggctacacgccggggtggagaccaccaccggtcgctcggccaggggctggccaacgccg
tktA2NotIB     1210  ------------------------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------------------------
tktAloppR       461  caagcgcttccgcgcagttgcattcgcgcactccgggcaccggagtacggctacacgccggggtggagaccaccaccggtcgctcggccaggggctggccaacgccg
Pm forw         582  caagcgcttccgcgcagttgcattcgcgcactccgggcaccggagtacggctacacgccggggtggagaccaccaccggtcgctcggccaggggctggccaacgccg
tktA24767         1  ------------------------------------------------------------------------------------------------------------
tktA12ATG       251  caagcgcttccgcgcagttgcattcgcgcactccgggcaccggagtacggctacacgccggggtggagaccaccaccggtcgctcggccaggggctggccaacgccg

Insert tktA2    396  tcggcatggccctggccgagcggctgctggcggcggcactccgggcgtgctgcggcggggcgacttcaaccgcgggacggatcgtcgaccacgccacctacgccttcgtcggcgacggctgcctgatggag
tktA2NotIB     1210  ------------------------------------------------------------------------------------------------------------
M13rev         1250  ------------------------------------------------------------------------------------------------------------
```

121

```
tktA1oppR      351  tcggcatggccctggccgagcggctgctggcggcggcggacttcaacgccgggacaggcgatcgtcgacaccgcacctacgccttcgtcgtcggcacggcgctgctgatggag
Pm forw        692  tcggcatggccctggccgagcggctgctggcggcggcggacttcaacgccgggacaggcgatcgtcgacaccgcacctacgccttcgtcgtcggcacggcgctgctgatggag
tktA24767      1    ------------------------------------------------------------------------------------------------------------------
tktA12ATG      361  tcggcatggccctggccgagcggctgctggcggcggcggacttcaacgccgggacaggcgatcgtcgacaccgcacctacgccttcgtcgtcggcacggcgctgctgatggag

Insert tktA2   506  ggcatctcccacgaggtctgctgctgccgccagcaccctggactcggcaagctgatctgttctacgacgacaacggcatctccatcgacggtgagtcgcggctgtt
tktA2NotIB     1177 gg--------------------------------------------------------------------------------------------------------
M13rev         1250 --------------------------------------------------------------------------------------------------------
tktA1oppR      241  ggcatctcccacgaggtctgctgctgccgccagcaccctggactcggcaagctgatctgttctacgacgacaacggcatctccatcgacggtgagtcgcggctgtt
Pm forw        802  ggcatctcccacgaggtctgctgctgccgccagcaccctggactcggcaagctgatctgttctacgacgacaacggcatctccatcga-ggtgagtcgcggctgtt
tktA24767      1    ---------------------------------------------------------------------------cggtgagtcgcggctg--t
tktA12ATG      471  ggcatctcccacgaggtctgctgctgccgccagcaccctggactcggcaagctgatctgttctacgacgacaacggcatctccatcgacggtgagtcgcggctgtt

Insert tktA2   616  cagcgacacgccggcgcggcgcttccgcgcctacggctggcaggtcgtcgaggcggtcgtcgaggcggtcgacggccacgacgcgcgctcgaggcggcgcgctgcgcggccaaggccg
tktA2NotIB     1175 -------------------------------------------------------------------------------------------------------------------
M13rev         1250 -------------------------------------------------------------------------------------------------------------------
tktA1oppR      131  cagcgacacgccggcgcggcgcttccgcgcctacggctggcaggtcgtcgaggcggtcgac-gccacgacgcgcgctcgaggcggcgcgctgcgcggccaaggccg
Pm forw        911  cagcgacacgccggcgcggcgcttccgcgcctac-gctggcaggtcgtcgaggcggtcgac-gccacgacgcgcgctcggg-gcggcgcgctcggcggcggcgctgcgcggccaaggccg
tktA24767      20   cagcgacacgccggcgcggcgcttccgcgcctacggctggcaggtcgtcgaggcggtcgtcgaggcggtcgacggccacgacgcgcgctcgaggcggcgcgctgcgcggccaaggccg
tktA12ATG      581  cagcgacacgccggcgcggcgcttccgcgcctacggctggcaggtcgtcgaggcggtcgtcgaggcggtcgacggccacgacgcgcgctcgaggcggcgcgctgcgcggccaaggccg

Insert tktA2   726  atacgagcggccgacgcgctgatctgctgcaagaccaccatcggcttcggcgtcgcgccgaccgcagccacgactgccacgcgcccgctgggcggccgaggagatc
tktA2NotIB     1175 ---------------------------------------------------------------------------------------------------------------
M13rev         1250 ---------------------------------------------------------------------------------------------------------------
tktA1oppR      23   ----------------------------------------------------------------------------------------------------------------
Pm forw        1017 ----------------------------------------------------------------------------------------------------------------
tktA24767      130  atacgagcggccgacgcgctgatctgctgcaagaccaccatcggcttcggcgtcgcgccgaccgcagccacgactgccacgcgcccgctgggcggccgaggagatc
tktA12ATG      691  atacgagcggccgacgcgctgatctgctgcaagaccaccatcggcttcggcgtcgcgccgaccgcagccacgactgccacgcgcccgctgggcggccgaggagatc

Insert tktA2   836  gccgcggccccgcgaggcgctggactgccgcgacgcgccctcgagctgccgcagatcgcccaggcctggacgcggcgcgagcgcggggcgctggagcgcgggactg
tktA2NotIB     1175 ----------------------------------------------------------------------------------------------------------------
M13rev         1250 ----------------------------------------------------------------------------------------------------------------
tktA1oppR      23   ----------------------------------------------------------------------------------------------------------------
Pm forw        1102 ----------------------------------------------------------------------------------------------------------------
tktA24767      240  gccgcggccccgcgaggcgctggactgccgcgacgcgccctcgagctgccgcagatcgcccaggcctggacgcggcgcgagcgcgggcgcgcgctggagcgcgggactg
tktA12ATG      801  gccgcggccccgcgaggcgctggactgccgcgacgcgccctcgagctgccgcagatcgcccaggcctggacgcggcgcgagcgcgggcgcgcgctggagcgcgggactg

Insert tktA2   946  gcgacaacgcttcgacgactacgcgc-ggcctaccggcgccctggccgcgccgagctgggccggacgacgcgc-ggcctaccggcgccctggccgcgccgactgggg-cgcgcggcgcgacg
tktA2NotIB     1139 gcgacaacgcttcgacgactacgcgc-ggcctaccggcgccctggccgcgccgagctgggccggacgacgcgc-ggccctaccggcgccctggccgcgccgactgggg-cgcgcggcgcgacg
M13rev         1228 ----------------------------------------------------------------------------------------------------------------
tktA1oppR      23   ----------------------------------------------------------------------------------------------------------------
Pm forw        1102 ----------------------------------------------------------------------------------------------------------------
tktA24767      350  gcgacaacgcttcgacgactacgcgc-ggcctaccggcgccctggccgcgccgagctggcggccgactggg-cgcgcggcgcgacg
tktA12ATG      911  gcgacaacgcttcgacgactacgcgc-ggcctaccggcgccctggccgcgccgagctgcgcggccgactgggg-cgcg-ggcgcgacg
```

122

```
Insert tktA2  1053  cgctgatcgcc-ggctgcgcagcaggcggcggcgagcagcagagctcgcgagtcctgaagcgctgggccgctgctgccgg-aactgctcggcggct
tktA2NotIB    1032  cgctgatcgcc-ggctgcgcagcaggcggcggcgagcagcagagctcgcgagtcctgaagcgctgggccgctgctgccgg-aactgctcggcggct
M13rev        1161  cgctgatcgcc-ggct-ccagcaggcggcggcgagcagcagagctcgcgagtc---gaagcgctgggccgctgctgcc-g-aaatgctcggccgct
tktAloppR       23  -----------------------------------------------------------------------------------------------
Pm forw       1102  -----------------------------------------------------------------------------------------------
tktA24767      457  cgctgatcgcc-ggctgcgcagcaggcggcggcgagcagcagagctcgcgagtcctgaagcgctgggcccgctgctgccgg-aactgctcggcggct
tktA12ATG     1018  -----------------------------------------------------------------------------------------------

Insert tktA2  1161  cggcggacctgcgccctcgaacctgaccgcgctggtccgctccgctcggtcggcggcggcgaggcgccggggcaactacatccactacgggtgcgcgagttcggcatg
tktA2NotIB     922  cggcggacctgcgccctcgaacctgaccgcgctggtccgctccgctcggtcggcggcggcgaggcgccggggcaactacatccactacgggtgcgcgagttcggcatg
20090223_A03  1058  cggcggacctgcgccctcgaacctgaccgcgctggtccgctccgctcggtcggcggcggcgag-cgcccggggcaactacatccactacgggtgcgcgagttcggcatg
tktAloppR       23  -----------------------------------------------------------------------------------------------------------
Pm forw       1102  -----------------------------------------------------------------------------------------------------------
tktA24767      565  cggcggacctgcgccctcgaacctgaccgcgctggtccgctccgctcggtcggcggcggcgaggcgccggggcaactacatccactacgggtgcgcgagttcggcatg
tktA12ATG     1073  -----------------------------------------------------------------------------------------------------------

Insert tktA2  1271  agcgcgatgatgaacggctggcgctgcacggcggtttcattcctacggcggcgcacttcctgatgttcatgaatacgcccgcaacgcgctgcatggccatggccgctgat
tktA2NotIB     812  agcgcgatgatgaacggctggcgctgcacggcggtttcattcctacggcggcgcacttcctgatgttcatgaatacgcccgcaacgcgctgcatggccatggccgctgat
M13rev         949  agcgcgatgatgaacggctggcgctgcacggcggtttcattcctacggcggcgcacttcctgatgttcatgaatacgcccgcaacgcgctgcatggccatggccgctgat
tktAloppR       23  ---------------------------------------------------------------------------------------------------------------
Pm forw       1128  ---------------------------------------------------------------------------------------------------------------
tktA24767      675  agcgcgatgatgaacggctggcgctgcacggcggtttcattcctacggcggcgcacttcctgatgttcatgaatacgcccgcaacgcgctgcatggccatggccgctgat
tktA12ATG     1100  ---------------------------------------------------------------------------------------------------------------

Insert tktA2  1381  gaagctgcgcgcggtgttcgtctacaccacgactcgatcggctggcgagacggccgaccccaccagcc-ggtcgaacagctcgcgcgctgcgcgccagcgccgaac
tktA2NotIB     702  gaagctgcgcgcggtgttcgtctacaccacgactcgatcggctggcgagacggccgaccccaccagcc-ggtcgaacagctcgcgcgctgcgcgccagcgccgaac
M13rev         839  gaagctgcgcgcggtgttcgtctacaccacgactcgatcggctggcgagacggccgaccccaccagcca-cctgcgcgaacagctcgcgcgctgcgcgccagcgccgaac
tktAloppR       23  -------------------------------------------------------------------------------------------------------------
Pm forw       1177  -------------------------------------------------------------------------------------------------------------
tktA24767      785  gaagctgcgcgcggtgttcgtctacaccacgactcgatcggctggcgagacggccgaccccaccagcccgtcgaacagctcgcgcgctgcgcgccagcgccgaac
tktA12ATG     1117  -------------------------------------------------------------------------------------------------------------

Insert tktA2  1490  ctggaaacctgcgcggcccctgcgacgagacgagaccgcgcggtggcctggtgcgcgggacggccgtgtcggcgctcgaacggcgggacggccgtgtgctgtcgcgcggcagaacctgcc
tktA2NotIB     593  ctggaaacctgcgcggcccctgcgacgagacgagaccgcgcggtggcctggtgcgcgggacggccgtgtcggcgctcgaacggcgggacggccgtgtgctgtcgcgcggcagaacctgcc
M13rev         731  ctggaaacctgcgcggcccctgcgacgagacgagaccgcgcggtggcctggtgcgcgggacggccgtgtcggcgctcgaacggcgggacggccgtgtgctgtcgcgcggcagaacctgcc
tktAloppR       23  -----------------------------------------------------------------------------------------------------------------------
Pm forw       1177  -----------------------------------------------------------------------------------------------------------------------
tktA24767      895  ctggaaacctgcgcggcccctgcgacgagacgagaccgcgcggtggcctggtgcgcgggccgt-acggcgggacggcccgt-ggcgctggtggtgctgtcgcgggcaga-------cc
tktA12ATG     1117  -----------------------------------------------------------------------------------------------------------------------

Insert tktA2  1600  ggccgcgcgcgcgacgcgcggcagcgcgcgcgcagcgcgcggcggcccatcgagcgcgcggcacgccggacgcgcgcgagctgatcctgctcgactgcggctcggcgctccgagg
tktA2NotIB     483  ggccgcgcgcgcgacgcgcggcagcgcgcgcgcagcgcgcggcggcccatcgagcgcgcggcacgccggacgcgcgcgagctgatcctgctcgactgcggctcggcgctccgagg
M13rev         621  ggccgcgcgcgcgacgcgcggcagcgcgcgcgcagcgcgcggcggcccatcgagcgcgcggcacgccggacgcgcgcgagctgatcctgctcgactgcggctcggcgctccgagg
```

123

```
tktAloppR     23  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767    998  tgccgcggcgcgcgaccaggcagcgcgcgg-catcgagcgcg-cgcctatgtgctgctcgactgccgggcac--------
tktA12ATG   1138  --------------------------------------------------

Insert tktA2 1710  tggcgctggcgctggcgcggcgggcaggcgctgggcgaggcggccggcggtccggtgtcgcgtgctgccgtgcgtttcgagcgcaggcggcggcggactaccgc
tktA2NotIB   373  tggcgctggcgctggcgcggcgggcaggcgctgggcgaggcggccggcggtccggtgtcgcgtgctgccgtgcgtttcgagcgcaggcggcggcggactaccgc
M13rev       511  tggcgctggcgctggcgcggcgggcaggcgctgggcgaggcggccggcggtccggtgtcgcgtgctgccgtgcgtttcgagcgcaggcggcggcggactaccgc
tktAloppR     23  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767   1071  --------------------------------------------------
tktA12ATG   1138  --------------------------------------------------

Insert tktA2 1820  gaggcggtgctgccgtccggtgcgggcccgcgtcgcggtcgagatggcgcgtccggagagctgcaccgctacgtggtctcgacggcggcggtgtctcggcatggaggg
tktA2NotIB   263  gaggcggtgctgccgtccggtgcgggcccgcgtcgcggtcgagatggcgcgtccggagagctgcaccgctacgtggtctcgacggcggcggtgtctcggcatgaggg
M13rev       401  gaggcggtgctgccgtccggtgcgggcccgcgtcgcggtcgagatggcgcgtccggagagctgcaccgctacgtggtctcgacggcggcggtgtctcggcatggaggg
tktAloppR     23  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767   1071  --------------------------------------------------
tktA12ATG   1204  --------------------------------------------------

Insert tktA2 1930  gttcggcgagtcggcgcggcggcgagctgttccggcatttcggctcagcgtcgagaacctgctcgccaggcgcaggccgtgctgccggcctgagggagggagggcgt
tktA2NotIB   153  gttcggcgagtcggcgcggcggcgagctgttccggcatttcggctcagcgtcgagaacctgctcgccaggcgcaggccgtgctgccggcctgagggagggagggcgt
M13rev       291  gttcggcgagtcggcgcggcggcgagctgttccggcatttcggctcagcgtcgagaacctgctcgccaggcgcaggccgtgctgccggcctgagggagggagggcgt
tktAloppR     23  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767   1109  --------------------------------------------------
tktA12ATG   1204  --------------------------------------------------

Insert tktA2 2040  agccatgtatctgggcgtcgattgcgggaccaaggtggtgatactcgatccggagcggccgcata--------
tktA2NotIB    43  --------------------------------------------------
M13rev       181  agccatgtatctgggcgtcgattgcgggaccaaggtggtgatactcgatccggagcggccgccatagtgtataagagtcagaattcctcgagacgtacggt
tktAloppR      0  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767   1137  --------------------------------------------------
tktA12ATG   1204  --------------------------------------------------

Insert tktA2 2114  --------------------------------------------------
tktA2NotIB     0  --------------------------------------------------
M13rev        71  --------------------------------------------------
tktAloppR      0  --------------------------------------------------
Pm forw     1210  --------------------------------------------------
tktA24767   1184  --------------------------------------------------
tktA12ATG   1204  --------------------------------------------------
```

124

H.3 Alignment of *opptktA3* insert and *opptktA3* designed in Clone Manager

The alignment consists of the following sequences (row labels, repeated across blocks):

- Insert opptktA3
- tktA3nedF
- M13rev
- tktA3nedR
- tktA24767

```
Insert opptktA3 1504 gtggcatcgagcgcctgcttccggaacttccgggcgcattcaggccaaagccgtgcggtgccgacgctcagacgctcgacatacctgcctcgacatcaacctgcagacctggccgaggagtcaaccgtgctgctgcgccagg
tktA3nedF      1060 ----------------------------------------------------------------------------------------------------------------------------------------------
M13rev          248 gtggcatcgagcgcctgcttccggaacttccgggcgcattcaggccaaagccgtgcggtgccgacgctcagacgctcgacatacctgcctcgacatcaacctgcagacctggccgaggagtcaaccgtgctgctgcgccagg
tktA3nedR       169 gtggcatcgagcgcctgcttccggaacttccgggcgcattcaggccaaagccgtgcggtgccgacgctcgacacgctcgacatcaacctgcctcgacatcaacctgcagacctggccgaggagtcaaccgtgctgctgcgccagg
tktA24767      1153 ----------------------------------------------------------------------------------------------------------------------------------------------

Insert opptktA3 1654 ctgccgggcaggggccgctcaagggattgctggactacaccgaactagt----------------
tktA3nedF      1060 ----------------------------------------------------------------
M13rev           98 ctgccgggcaggggccgctcaagggattgctggactacaccgaactagt----------------
tktA3nedR        19 ----------------------------------------------------------------
tktA24767      1205 ----------------------------------------------------------------
```

H.4 Alignment of *opptktA2* insert and *opptktA2* designed in Clone Manager

```
opptktA2    600  ----------------------------------------------------------------------------------
opptktA2    397  ----------------------------------------------------------------------------------

Insert      811  atcaactgaaactggaggcgc---tgctgcgccgactcgccggaggtgtggcatg-acccgaacctgaaacccggccgacggccgacctggccaatgc-catccgctcctgagcatggatgcggtggaacgggcgaactccggccat
opptktA2    455  ----------------------------------------------------------------------------------
opptktA2    121  atcaactgaaactggaggcgc---tgctgcgccgactcgccggaggtgtggcatg-acccgaacctgaaacccggccgacggcccggccctggccaatgc-catccgctcctgagcatggatgc-----------------------
opptktA2    355  ----------------------------------------------------------------------------------
opptktA2    874  atcaactgaaactgg-gacgc---tgctgcgccccgactcgccggaggtgtgggcatg-acccga--------------------
opptktA2     39  ----------------------------------------------------------------------------------
opptktA2     21  ----------------------------------------------------------------------------------
opptktA2    600  ------------------------------------------------------------ggcaatgc-catccgctcctgagcatgatgcggtggaacggggcgactccggccat
opptktA2    397  ----------------------------------------------------------------------------------

Insert      956  ccggcgccccatgggcatggccgatatcgccgagtgctctgaacggccacctgccacaaccccggtgaaccgccgcttcgtgtctcacacggcacggctcgatgctgctctacgcctgctgcac
opptktA2    455  ----------------------------------------------------------------------------------
opptktA2      0  ----------------------------------------------------------------------------------
opptktA2    355  ----------------------------------------------------------------------------------
opptktA2    939  ----------------------------------------------------------------------------------
opptktA2     39  ----------------------------------------------------------------------------------
opptktA2     21  ----------------------------------------------------------------------------------
opptktA2    543  ccggcgcccatggcatggccgatatcgccgagtgctctgaacggccaacccgcgccaaccccggccgacggctcgatgctgctctacgcctgctgcac
opptktA2    397  ----------------------------------------------------------------------------------

Insert     1106  ctctccggctacgacttgcggtcgagagctcaagcgcttccgccagttgcattcgcgactcccggcaccggtacacgccgggggtcggccagggctggccaacgccgtcggcatg
opptktA2    455  ----------------------------------------------------------------------------------
opptktA2      0  ----------------------------------------------------------------------------------
opptktA2    355  ----------------------------------------------------------------gggcaccggagtacggctacacgccgggggtggagcaccaccggtcgttcggccaacgccgtcggcatg
opptktA2    939  ----------------------------------------------------------------------------------
opptktA2     39  ----------------------------------------------------------------------------------
opptktA2     21  ----------------------------------------------------------------------------------
opptktA2    393  ctctccggctacgacttgcggtcgaggagctcaagcgcttccgccagttgcattcgcgactcccggcaccggagtacggctacacgccggggtggagcaccaccggtcggccaacgccgtcggcatg
opptktA2    397  ----------------------------------------------------------------------------------

Insert     1256  gccctggccgagcggcgtgctggcgggcgggcgacttcaaccgccggacaggcgatcgtgacaccgcacctaggcttcgtcgtcggcgggcgatctcccagaggtctgctgctcgtggcgcaacctgggactcggc
opptktA2    455  ----------------------------------------------------------------------------------
opptktA2      0  ----------------------------------------------------------------------------------
opptktA2    271  gccctggccgagcggcgtgctggcgcggcggcgacttcaaccgccggacaggcgatcgtgacaccgcacctaggcttcgtcgtcggcgggcgatctcccagaggtctgctgctcgtggcgcaacctgggactcggc
opptktA2    939  ----------------------------------------------------------------------------------accctgg---
opptktA2     39  ----------------------------------------------------------------------------------
opptktA2     21  ----------------------------------------------------------------------------------
opptktA2    243  gccctggccgagcggcgtggcgggcgcgcgacttcaaccgccggacaggcgatcgtgacaccgcacctaggcttcgtcgtcggcgggcgatctcccagaggtctgctgctcgtggcgcaacctgggactcggc
opptktA2    397  ----------------------------------------------------------------------------------

Insert     1406  aagctgatcgttctacgacgacaaacggcatccat-------------------------------
opptktA2    455  ----------------------------------------------------------------------------------
opptktA2      0  ----------------------------------------------------------------------------------
opptktA2    121  aagctgatcgtgttctacgacgacaacggcatccatcacctccatatcgaattcctgcagcccgggggatccacgagttc-----
opptktA2    946  ----------------------------------------------------------------------------------
opptktA2     39  ----------------------------------------------------------------------------------
opptktA2     21  ----------------------------------------------------------------------------------
opptktA2     93  aagctgatcgtgttctacgacgacaacggcatccatatc-------------------------
opptktA2    397  ----------------------------------------------------------------------------------
```

Appendix I The kinetics activity of TktA1, TktA2 and TktA3.

The measured activity of the three transketolase proteins in the cell free extract using xylulose-5-phosphate and ribose 5-phosphate as substrates. The activity was determined by measuring spectrophotometrically the increase in NADH absorbance at 340 nm and 37°C temperature. The protein samples are measured in triplicate.
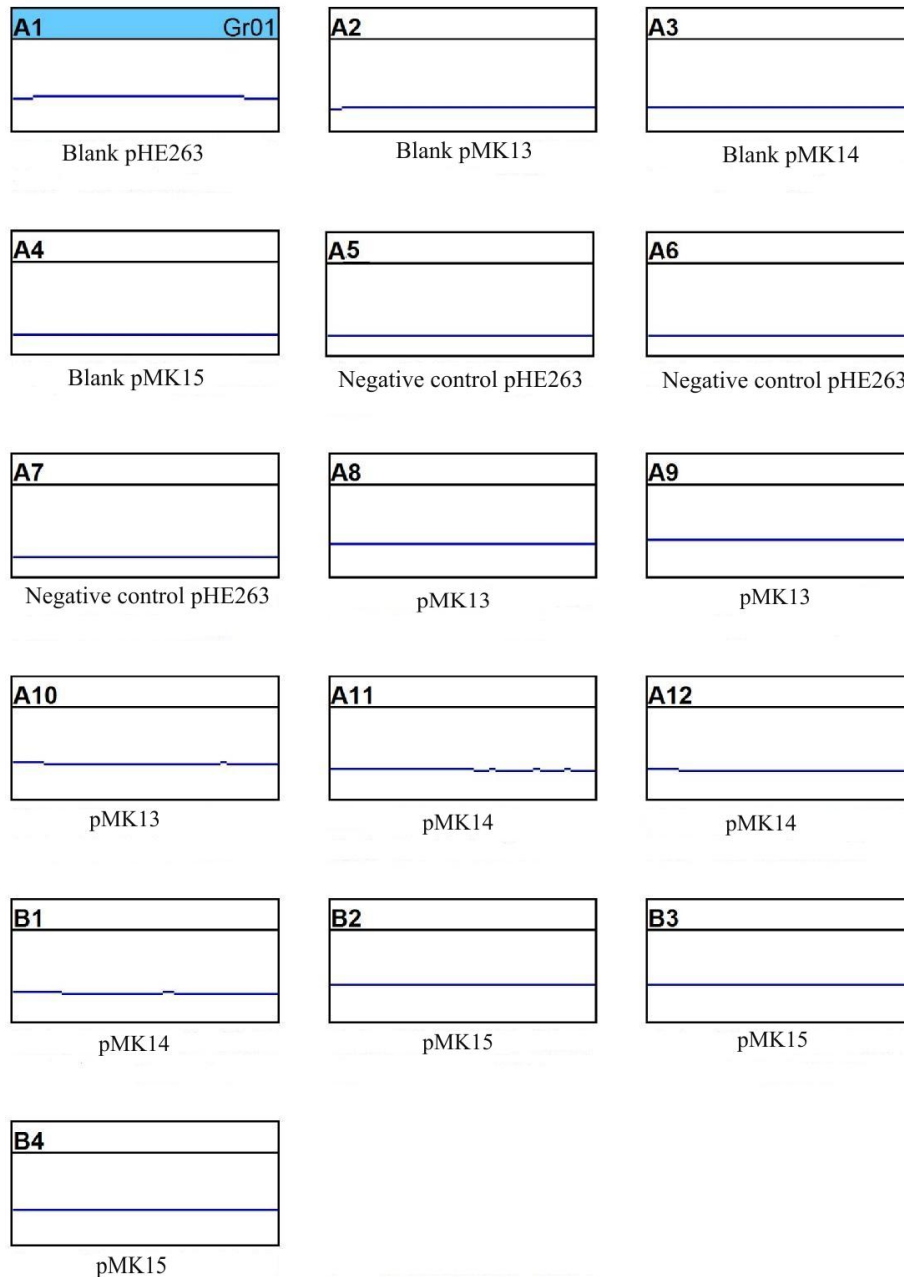
Figure I, The spectrophotometricallt activity measurements for the three transketolase genes in E. coli. *The blank samples (A1-A4) are with no added enzyme, samples A5-A7 are negative controls of pHE263 plasmid do not contain transketolase, samples A8-A10 for pMK13 containing tktA3, A11, A12 and B1 for pMK14 containing tktA1 and B2-B4 are for pMK15 containing tktA2.*