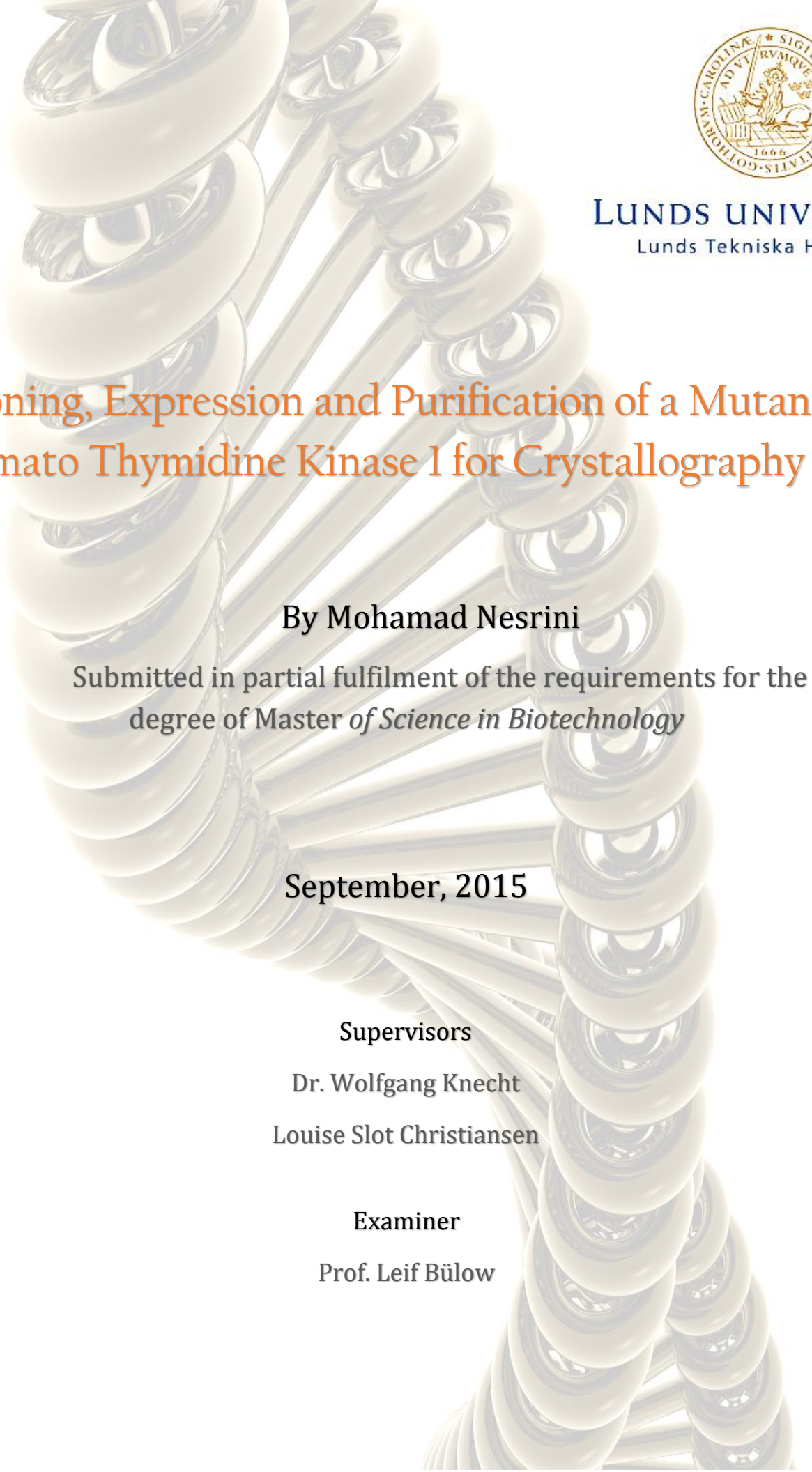

LUNDS UNIVERSITET
Lunds Tekniska Högskola

Cloning, Expression and Purification of a Mutant Tomato Thymidine Kinase 1 for Crystallography

By Mohamad Nesrini

Submitted in partial fulfilment of the requirements for the degree of Master *of Science in Biotechnology*

September, 2015

Supervisors

Dr. Wolfgang Knecht

Louise Slot Christiansen

Examiner

Prof. Leif Bülow

Contents

Abstract

Thymidine kinase 1 from tomato (ToTK1) is a newly discovered deoxyribonucleoside kinase (dNK) that has been subject to study due to its *in vitro* and *in vivo* properties making it an interesting suicide gene candidate for the treatment of brain tumors. The ToTK1 phosphorylates the nucleoside analogue 3-azido-2,3-dideoxythymidine (azidothymidine, AZT) equally well as its natural substrate thymidine (dThd). A truncated ToTK1 showed improved activity twice as high as the wild-type ToTK1, and also an increased specificity towards AZT. The combination of ToTK1 and AZT has been previously tested in two animal studies for its efficiency and use in suicide gene therapy for malignant glioma. In the present study, the C-terminal truncated version of ToTK1 (ToTK1ΔC25) has been successfully expressed in *E. coli* and purified by means of affinity chromatography (IMAC). Although, the purified ToTK1ΔC25 of bacterial origin displayed no significant enzymatic activity, expression was attempted in two insect cell lines, *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Hi5) using the baculovirus expression vector system (BEVS). Significant TK activities were detected in the crude extracts of insect cells transfected with recombinant baculoviruses bearing N- and C-terminal His-tagged ToTK1ΔC25 constructs. The findings of this project pave the way to determine the 3D structure of ToTK1 for understanding the structure-function relationships of nucleoside activation by this enzyme and thereby show routes towards further improvement of ToTK1 and other TK1-like dNKs for use in suicide gene therapy.

Key words: nucleosides, nucleoside analog, deoxyribonucleoside kinases, thymidine kinase, Azidothymidine, protein engineering, protein expression, suicide gene therapy, BEVS.

Abbreviations

Enzymes

dNK	Deoxyribonucleoside kinases
RNR	Ribonucleotide Reductase
TK	Thymidine kinases
Wt-ToTK1	Wilde-type Tomato thymidine kinase 1
ToTK1ΔC25	C-terminal truncated Tomato thymidine kinase 1 (25 amino-acid deletion)
N-His	N-terminal histidine-tag
C-His	C-terminal histidine-tag
hTK1	Human thymidine kinase 1
HSV-TK	Herpes simplex virus thymidine kinase
Dm-dNK	*Drosophila melanogaster* dNK
dCK	Deoxycytidine kinase
dGK	Deoxyguanosine kinase
TEV	Tobacco Etch Virus Protease

Nucleosides/Nucleotides and Nucleoside Analogues

dN	Deoxyribonucleoside
dAdo	Deoxyadenosine
dThd	Deoxythymidine
dUrd	Deoxyuridine
dCyd	Deoxycytidine
dGuo	Deoxyguanosine
-MP	Monophosphate
-DP	Diphosphate
-TP	Triphosphate
NDPs	Nucleoside Diphosphates
NA	Nucleoside analogue
AZT	Azidothymidine (3'-azido-2', 3'-dideoxythymidine)
GCV	Ganciclovir (9-(1, 3-dihydroxy-2-propoxymethyl)-guanine)

Other

GDEPT	Gene-Directed Enzyme Prodrug Therapy
BEVS	Baculovirus Expression Vector System
MOI	Multiplicity of Infection
LP3	Lund Protein Production Platform, Lund University

Introduction

DNA biosynthesis

Deoxyribonucleoside triphosphates (dNTPs) are the building blocks of DNA in living organisms. Balanced and constant supply of all four dNTPs is crucial for nucleic acid synthesis and repair. The precursors for the four deoxyribonucleoside 5´-triphosphates dATP, dCTP, dGTP and dTTP (dNTPs) are produced via two pathways, the *de novo* and the *salvage* pathway (Fig. 1). (Buchanan and Hartman 1959, Thelander and Reichard 1979, Arnér and Eriksson 1995).

The *de novo* pathway is a major source for dNTP synthesis in replicating cells. This metabolic pathway utilizes basic molecules such as: sugars, amino acids, CO_2, NH_3, etc. to build dNTPs (Zhang 2008). The *de novo* pathway is controlled by ribonucleotide reductase (RNR), which catalyzes the reduction of ribonucleoside diphosphates (NDPs) to deoxyribonucleoside diphosphates (dNDPs) by utilizing NADPH as an electron donor, and produces the precursors for DNA synthesis. RNR is considered the key enzyme that plays a crucial role in regulation of dNTPs generation for DNA synthesis (Thelander and Reichard, 1979).

On the other hand, the *salvage* pathway is an important alternative source that also supplies dNTPs. Deoxyribonucleosides (dNs) formed from degradation of DNA can be recycled by the salvage pathway and used for DNA repair and mtDNA synthesis in resting cells since there is no *de novo* dNTP synthesis in these cells (Yi-Zheng Xu et al., 1995). The dNs are transported into the cell by nucleoside transporter proteins and phosphorylated to their respective monophosphates by enzymes called deoxyribonucleoside kinases (dNKs). dNKs are the key enzymes in the *salvage* pathway, which catalyze the transfer of a phosphoryl group from ATP to a deoxyribonucleoside, thereby forming deoxyribonucleoside monophosphates (dNMPs). The dNMPs are further phosphorylated by nucleoside monophosphate kinases (NMPKs) to their corresponding diphosphates, and finally to dNTPs by nucleoside diphosphate kinases (NDPKs) (Reichard 1988, Arnér and Eriksson 1995).

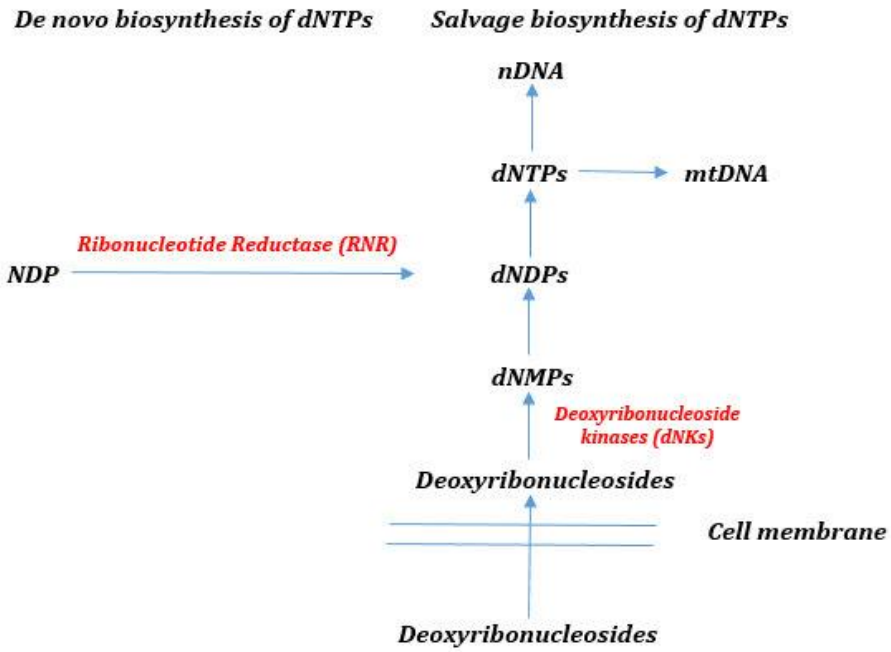

Figure 1 Deoxyribonucleoside 5´-triphosphates are synthesized via *de novo* and *salvage* pathways and used for the synthesis of mitochondrial DNA (mtDNA) and nuclear DNA (nDNA).

Deoxyribonucleoside kinases (dNKs)

dNKs are found in all studied organisms, except fungi, and exhibit a high level of diversity among the analyzed living organisms. And as mentioned previously, dNKs catalyse the rate-limiting step of the *salvage* pathway by phosphorylation of deoxyribonucleosides and their analogues to their corresponding monophosphate compounds (Christiansen, 2015).

Based on their three-dimensional structures and amino acid sequence similarities, dNKs can be broadly devided in two large groups: the Thymidine Kinase 1 (TK1)-like kinases and the non-TK1-like kinases (Clausen, 2012; Sandrini and Piskur, 2005). In human cells, there are four dNKs: deoxyguanosine kinase (dGK), deoxycytidine kinase (dCK), thymidine kinase 1 (TK1) and thymidine kinase 2 (TK2). The mammalian deoxyribonucleoside kinases show distinct but partially overlapping substrate specificities (**Table 1**) (Arnér and Eriksson 1995). The two thymidine kinases, TK1 and TK2 are strict pyrimidine kinases; TK1 only phosphorylates deoxythymidine (dThd) and deoxyuridine (dUrd), whereas the mitochondrial TK2 in addition to dThd and dUrd also phosphorylates deoxycytidine (dCyd). Deoxycytidine kinase has broad substrate specificity and phosphorylates dCyd as well as deoxyadenosine (dAdo) and deoxyguanosine (dGuo). Finally, dGK, a purine kinase, can phosphorylate dAdo and dGuo. TK2 and dGK are found in mitochondria while dCK and TK1 are cytosolic. Both compartments are therefore able to salvage all deoxyribonucleosides [dThd, (and dUrd), dCyd, dGuo and dAdo] needed for DNA synthesis (Arnér and Eriksson 1995).

Table 1 Substrate specificity and localization of deoxyribonucleoside kinases.

DNKs	Substrate	Localization
TK1	dThd , dUrd	Cytosol
TK2	dThd, dUrd, dCyd	Mitochondria
dCK	dCyd, dAdo, dGuo	Cytosol
dGK	dAdo, dGuo	Mitochondria

Biodiversity of dNKs

The biodiversity of dNKs has been recently reviewed by Christiansen et al. (Christiansen, 2015). In addition to mammals, dNKs have been studied and characterized in other vertebrates such as fish, frog (*Xenopus laevis* and *Xenopus tropicalis*) and bird (*Gallus gallus*) (Knecht et al., 2002a; Mutahir et al., 2013). However, a duplicated homolog of dCK has been found in those vertebrates, while *G. gallus* has lost its dGK and possibly replaced it by the duplicated dCK (Konrad et al., 2014).

Insect dNKs

dNKs have been also characterized from insects like, *Drosophila melanogaster* and *Anopheles gambiae*. The insect dNK is a non-TK1 like kinase and it exists as individual dNK, which has all four dNK activities (Knecht et al., 2002a; Knecht et al., 2003; Munch-Petersen et al., 2000; Munch-Petersen et al., 1998a, b). The multisubstrate kinase from *Drosophila melanogaster* (DmdNK) is considered as the fastest dNK characterized so far due to its high $k_{cat,}$ broad substrate specificity and fast turnover. Those important properties have made DmdNK an interesting candidate for some biotechnological and medical applications. Especially, for suicide gene therapy (Christiansen et al., 2015).

Bacterial dNKs

Different bacterial dNKs have been characterized from both Gram-negative bacteria like *E. coli* and Gram-positive bacteria like *Bacillus* species. Bacterial dNKs are generally considered as non-TK1-like kinases (Clausen et al., 2012), and it has been thought that Gram-negative bacteria have only one dNK, TK1, while Gram-positive bacteria seem to have several dNKs (Sandrini et al, 2007a, b). Moreover, dNKs from two aquatic bacteria, Gram-negative *Flavobacterium psychrophilum* JIP02/86 and *Polaribacter sp*. MED152, have been investigated by Tinta et al. (Tinta, 2012). The study has shown that several dNKs have been found including TK1-like and non-TK1-like kinases. Recently, a thymidine kinase from the protozoan parasite *Leishmania major* (Timm, 2015) has been characterized. Bacterial dNKs have been tested for their potential for suicide gene therapy (Hebrard et al., 2009).

Plant dNKs

A number of dNKs have been characterized from plant species such as *Arabidopsis thaliana* and *Solanum lycopersicum* (Khan, 2010). Three different dNKs from *A. thaliana* were characterized by (Clausen et al., 2012). Two of them are duplicated TK1 copies that

phosphorylate dThd and dUrd, but no other natural deoxynucleosides. The third dNK phosphorylates all natural substrates except dThd.

The Kinase activities were measured in *A. thaliana* cytosol, chloroplast and mitochondria. However, those kinase activities were mainly located in mitochondria, and no activities were detected in chloroplasts (Clausen et al., 2014). Recently, also the thymidine kinase called AtTK2 has been identified in the mitochondria of *A. thaliana* (Xu, 2015). Deoxycytidine (dCK) and deoxyguanosine kinases (dGK) have also been found in tomato, maize (*Zea mays*), pine (*Pinus taeda*) and rice (*Oryza sativa*) (Knecht, 2010).

Thymidine kinase 1 from tomato (ToTK1) was characterized in 2010 by Khan et al. and it has shown remarkable *in vitro* kinetic properties that made it a particular interesting candidate for suicide gene therapy (Khan et al., 2010).

Structural studies of dNKs

The structures of several dNKs have been solved including TK1-like kinases such as: human TK1, *Ureaplasma urealyticum* TK1, *Bacillus cereus* TK1, *Bacillus anthracis* TK1, and *Thermotoga maritime* TK1. And non-TK1-like kinases such as: human dCK, human dGK and the multisubstrate kinase from *Drosophila melanogaster* , DmNK. Recently, the 3D structure of Thymidine kinases from *Leishmania spp.(LmTK)* has been resolved and characterized by Timm et al. (Timm, 2015).

TK1-like kinases crystallize as tetramers consisting of two dimeric proteins that have distinct protein-protein interfaces. Each TK1 monomer contains two major domains with the active site buried in between them. The larger α/β domain is similar to the ATP-binding domain of the RecA-F1ATPase structural family. This domain has a six-stranded parallel β-sheet situated between a long α-helix and a flexible loop on one side and three shorter helices on the other side. The phosphate binding P-loop is highly conserved and it is situated in the junction between β1 and α1. This P-loop is also present in non-TK1-like kinases and it is similar in all dNKs. The other domain is a small domain, which is unique for TK1-like kinases, containing a structural zinc and a unique lasso-like loop that covers the phosphate acceptor site (Welin, 2004).

The substrate and the base bind in a hydrophobic cleft between the α/β domain and the lasso domain while the triphosphate binds to the α/β domain. The thymine base is hydrogen-bonded to the main-chain atoms only in a very narrow substrate pocket, this is most likely the reason for the restricted substrate specificity of the TK1 like kinases. On the other hand, the thymine base in the non-TK1-like dNKs is stabilized by different side chains that determine the different substrate specificities giving non-TK1-like dNKs special advantage to adapt to different substrates by few amino acid changes (Knecht et al., 2002b; Skovgaard et al., 2012).

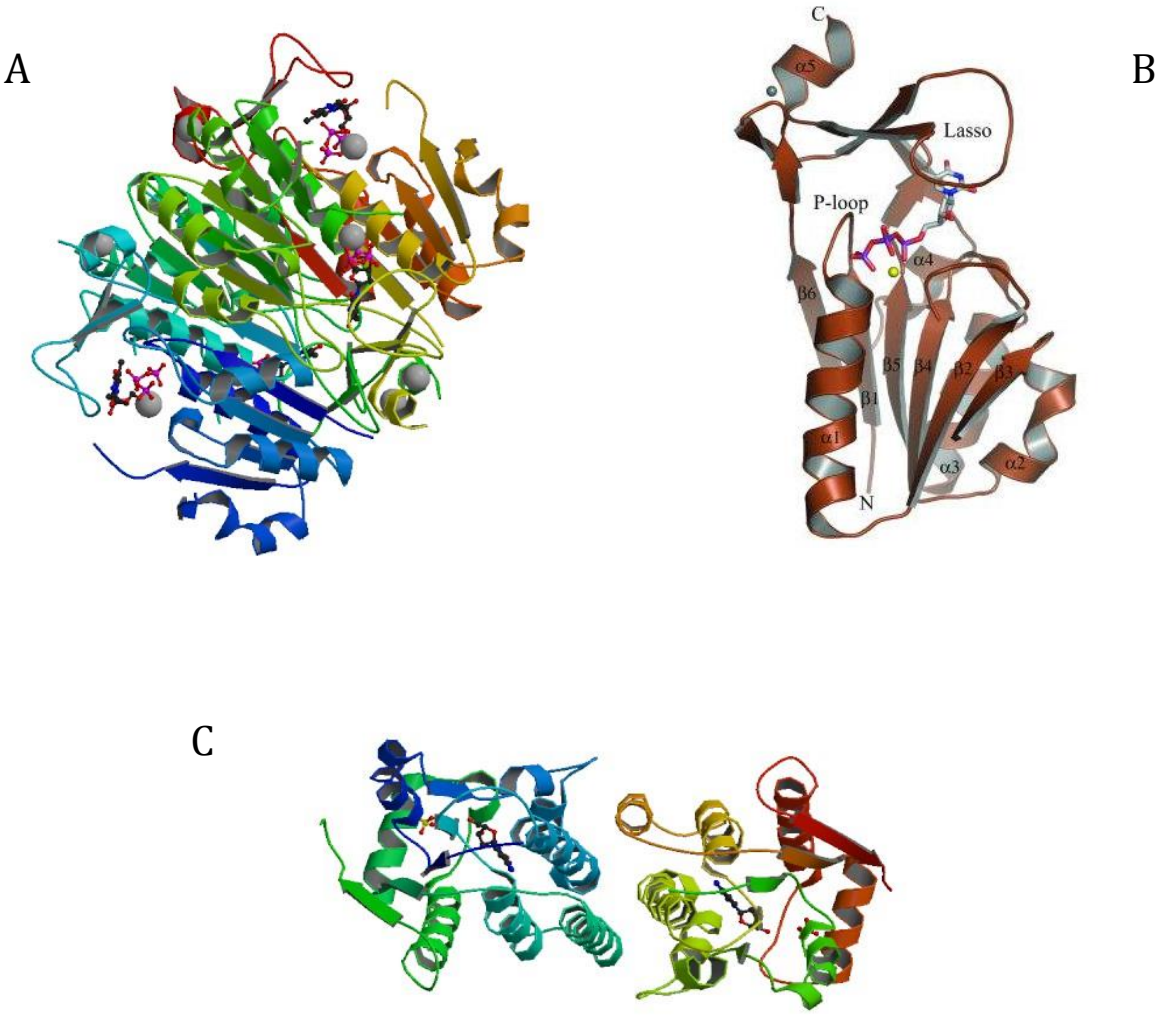

Figure 2. Three examples of 3D structure of Thymidine kinases retrieved from PDB:
A) Crystal Structure of Human Thymidine Kinase 1 (tetramer, **PDB_1XBT**)
B) Monomer structure of human TK1 with dTTP bound in the substrate binding site (Welin, 2004)
C) Crystal Structure of Drosophila Deoxyribonucleoside Kinase (dimer, **PDB_1J90**)
Protein chains are colored from the N-terminal to the C-terminal using a rainbow (spectral) color gradient.

Applications of dNKs

Activation of Nucleoside Analogues (NAs) by dNKs

Nucleoside analogues (NAs) are synthetic nucleosides that contain a modified base and/or sugar moiety. NAs work as antimetabolites by being similar enough to nucleosides to be phosphorylated by the enzymes of the salvage pathway (dNKs) and incorporated into growing DNA strands (Christiansen, 2015). NAs are clinically important compounds as they can be used as therapeutic drugs, in particular as anticancer and antiviral drugs.

The mechanism in which NAs are activated has been studied extensively in cancer research (Young et al, 2013). First, NAs enter the cells by nucleoside transporters (NTs) in the plasma membrane. Once inside the cell, the NAs are phosphorylated by the cells own dNKs or by viral dNKs to their monophosphate forms which are further phosphorylated to their active di- and triphosphate forms by NMPKs and NDPKs respectively (Young et al., 2013). The toxic effect of the activated NAs is executed by incorporation into elongating DNA causing chain termination and also by inhibition of intracellular enzymes, such as viral or human polymerases or RNR (Jordheim, 2013). (Fig. 3)

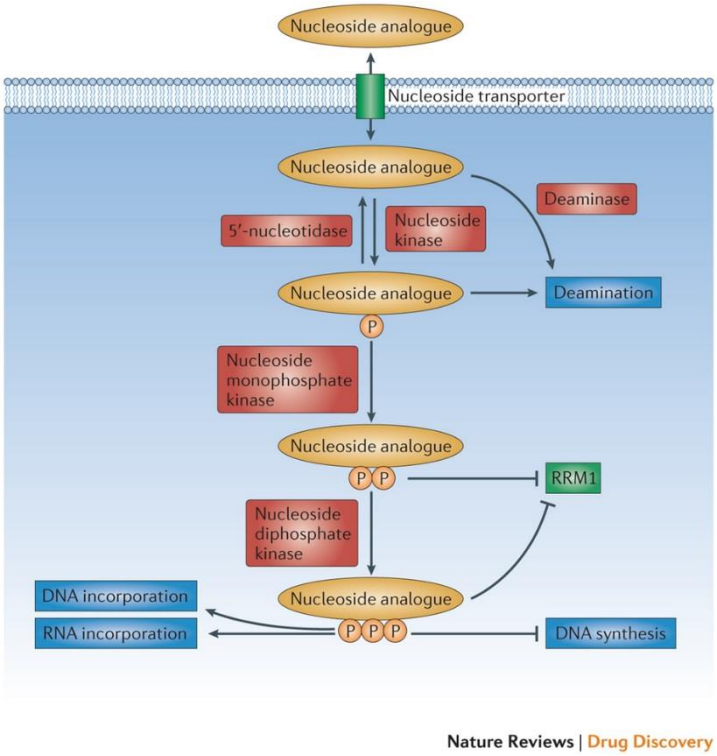

Figure 3. **Mechanism of action of nucleoside analogues (NAs). (Jordheim, 2013)**

Suicide Gene Therapy using dNKs

Conventional chemotherapies and surgical removal of tumor are the most common treatments of cancer. However, in many cases, e.g. brain tumors, the tumor is too complicated to remove by conventional means, and such therapies become ineffective. Moreover, chemotherapies have side effects on normal tissues (off target).

One of the most innovative approaches to developing antineoplastic agents with increased tumor selectivity is the use of suicide gene therapy (Rajab, 2013), which is also known as Gene-Directed Enzyme Prodrug Therapy (GDEPT). Suicide gene therapy aims at delivering a gene product into the targeted cancer tissue followed by tissue/tumor-specific expression of the gene product which results in an enzyme that has the ability to convert a non-toxic prodrug into a cytotoxic compound that terminates DNA elongation and causes cell apoptosis of the cancer cells (Aghi, 2000).

When the prodrug is administered, it will only be activated in the tumor cells where the suicide gene is expressed (Fig. 4). Although not all tumor cells will express the suicide gene, but due to the bystander effect, the cytotoxic effect spreads to the neighboring tumor cells resulting in apoptosis and regression of the whole tumor (freeman, 1993; Huber, 1994; Aghi, 2000).

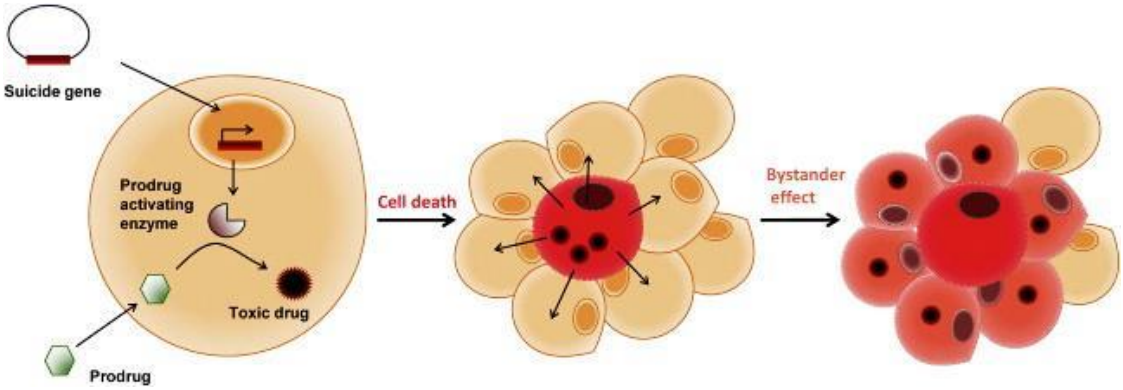

Figure 4 **General principle of a suicide gene/prodrug system.** In case of dNKs, the gene coding for a dNK is delivered to the target cell. The cellular expression of the dNK (prodrug activation enzyme) allows a nucleoside analog (prodrug) to be activated in the cell to its toxic form and subsequently kill the target cell (red). The toxic effect is transferred into the neighboring cells (arrows). This bystander effect allows the cytotoxc effect of the activated drug to spread to the neighboring cells and eradicate them, too (light red). **(Christiansen et al., 2015)**

The success of a suicide gene/prodrug system is determined by several important factors First of all, the suicide gene should be specifically expressed at high levels in the targeted tumor tissue. This is dependent on the delivery vector or vehicle, most often a virus. Second, the suicide enzyme has to have a high catalytic efficiency with the prodrug to effectively activate it. Third, the prodrug should not be activated by any other enzymes and induce toxicity in normal cells. And finally, a strong bystander effect should occur (Adriani et al., 2012).

Malignant Gliomas (MGs) are the most common primary malignant brain tumors (Omuro and DeAngelis 2013, Louis 2006) and they have a very poor prognosis and a significant decrease in quality of life. The fact that MGs consist of rapidly dividing cells confined to the central nervous system (CNS) in a specific location (brain), makes it very difficult to treat by conventional means such as chemotherapies. However, alternative and promising approaches such as gene therapy have recently been explored (Lawler 2006, Maatta 2009). The herpes simplex virus thymidine kinase (HSV-TK) with the prodrug ganciclovir (GCV) is the first well-characterized suicide gene/prodrug system that used for cancer therapy (Moolten, 1986; Osaki et al., 1994). (HSV-TK/GCV) system have been used in clinical gene therapy trials which proved its antitumor activity against different types of cancers, including malignant gliomas, leukemia, bladder cancer, liver cancer and head and neck tumors.

Although the HSV-TK/GCV system has been successfully applied in several models and showed good results, this system has disadvantages and limitations that reduce its treatment efficacy, such as the low activity of HSV-TK1, the lipophobicity of the prodrug GCV which limits its ability to pass the blood-brain barrier, and significant side effects have been also observed (Maatta 2009, Fillat 2003, Moolten and Wells 1990). Therefore, alternative suicide gene/prodrug system is needed as not all cancers are equally responsive to the same treatment.

Tomato Thymidine Kinase 1 (ToTK1) is a promising suicide gene system to be used in combination with the nucleoside analogue azidothymidine (AZT) as a prodrug. ToTK1 phosphorylates AZT, a drug originally used to treat HIV patients, equally well as its natural substrate dThd. In addition, it can phosphorylate dThd monophosphate and AZT-monophosphate to their diphosphate compounds, a property that not reported in any other dNKs (khan et al., 2010).

ToTK1/AZT system has been shown an efficient suicide gene therapy in malignant glioma cells (Khan, 2010). The antiretroviral prodrug used in this system, AZT, has several advantages that made it an ideal prodrug for the suicide gene therapy of brain tumors. AZT exhibits pharmacokinetic properties superior to other nucleoside analogs with better solubility. Moreover, AZT is more lipophilic than GCV and therefore it can pass the blood brain barrier more easily and effectively than GCV, this gives better access to the tumor in the brain (Khan 2010).

A study carried out by Khan et al. in 2010 has been shown that transduction of human glioblastoma cells with ToTK1 increases their sensitivity towards AZT by a factor of 500, which is superior to the effect produced by HSV-TK/GCV system. Moreover, in a mouse model using neural progenitor cells as delivery vehicle for ToTK1, a decrease in tumor size and increase in survival was observed in the mouse treated with AZT (Khan et al., 2010). The ToTK1/AZT system was recently evaluated as an alternative to HSV-TK/GCV for the treatment of malignant gliomas. Both treatments were effective at killing human MG cells *in vitro* and at decreasing tumor growth and improving survival of mice with malignant gliomas without significant differences (Stedt, 2015).

Because of the promising results of using ToTK1/AZT system in suicide gene therapy, further studies that aim at understanding of ToTK1 structure function relationship would be desirable. A recent study carried out by Christiansen et al. in 2015 has been shown that the catalytic efficacy for ToTK1 could be improved for AZT over dThd by random protein engineering and structural changes in ToTK1 that improves its efficiency as suicide gene in combination with AZT. A series of mutants of ToTK1 were generated by means of random mutagenesis by error-prone PCR as well as N- and C-terminal truncations (Christiansen 2015). Mutants were selected by their ability to sensitize *E. coli* KY895 toward AZT.

The mutant ToTK1ΔC25 was among the best performing mutants and it was able to decrease the minimum inhibitory concentration (MIC) of AZT compared with wild-type

ToTK1. The mutant ToTK1ΔC25 has a 25 amino acid deletion at the C-terminus, which has been shown to stabilize the enzyme *Dm*-dNK (Munch-Petersen 2000).

Kinetic studies showed an increase in specificity towards AZT as well as a significantly lower degree of inhibition by dTTP compared to the wild-type enzyme (Christiansen 2015). However, the improved kinetic properties of the mutants could not be correlated to specific amino acid residues. Therefore, a determination of the crystal structure of ToTK1 and its mutants would be necessary.

The Baculovirus Expression Vector System (BEVS)

The BEVS is one of the most powerful and widely used eukaryotic expression systems available. The BEVS is based on baculoviruses, which infect insect cells only as their natural hosts. Baculovirus has a large circular double-stranded DNA (between 80 and 200 kbp), a property that makes it possible to accommodate large amounts of foreign DNA by homologous recombination and generate recombinant particles. Therefore, baculoviruses are considered as highly interesting and efficient vectors to be used for heterologous gene expression (van Oers 2011). The most widely used baculovirus for recombinant protein expression is Autographa californica multiple nucleopolyhedrovirus (AcMNPV), which has a large circular dsDNA genome (130 kbp) (Kidd and Emery 1993, van Oers 2011). When compared with prokaryotic hosts such as E. coli, recombinant protein production using the BEVS approach has several advantages such as: multiple gene expression, the ability to produce high levels of recombinant protein with the appropriate post-translational modifications, the limited host range of baculoviruses, the lack of human pathogens in lepidopteran insect cells, the ability of insect cells to grow in serum-free media and without the need for carbon dioxide supplementation, and the ease of scale up (Kidd and Emery 1993, van Oers 2011).

Project goals

The aim of this degree project is to develop protocols for the recombinant production (molecular cloning and protein expression) of a mutant Tomato Thymidine Kinase 1 ToTK1ΔC25**,** which has a 25 amino acid deletion at the C-terminus. The recombinant production of ToTK1ΔC25 is aimed towards achieving high purity, yield and enzymatic activity of the produced enzyme that is suitable for crystallization experiments. The outcomes of this project will provide valuable information needed to resolve the 3D structures of ToTK1ΔC25. That will give better understanding of structure-function relationship of ToTK1.

To achieve this goal, two different expression hosts will be used: bacterial cells (*E. coli*) and insect cells (BEVS). Both expression systems will be compared to find out which of them is more suitable for protein expression based on several parameters such as protein yield, purity, titer, productivity, stability and enzymatic activity.

Materials

Please refer to **Appendix I. Materials** for a comprehensive list of chemical reagents, biological materials and equipment used to perform the experiments in this project.

Methods

Molecular Cloning of the mutant ToTK1 in *E. coli*

Transformation of *E. coli* TOP10 Competent Cells

The aim of this step was to transform 50 µL of One Shot TOP10® *E. coli* competent cells with ~100 ng of the intact plasmid vector pET-24(+) by using the heat shock method according to the instructions provided. Transformants were selected on LB agar plates containing 50 µg/mL kanamycin. Individual colonies were inoculated in 5 mL LB media containing kanamycin, and incubated with shaking overnight at 37 °C. Cells from the culture were harvested by centrifugation (8000 rpm, 2 min, RT) and the plasmids were isolated and purified using GeneJET Plasmid Miniprep Kit following the protocol provided.

Expression plasmid for the C-terminal His-tagged ToTK1ΔC25

The sequence encoding the toTK1 with a 25 amino acid deletion at the C-terminus was tagged with polyhistidine (8×His) and codon-optimized for expression in *Homo sapiens.* This construct was named ToTK1ΔC25-His to refer to the polyhistidine tag at the C-terminus. The ToTK1ΔC25-His was obtained from previously constructed pDual-HoHis2 expression vector that was kindly provided by the supervisor (L.S.C). Both vectors, pET-24a(+) and pDual-HoHis2, were digested at equivalent amounts (1 µg each) with 1µL of FastDigest™ enzymes *Nde I* and *Not I* following the protocol provided. The digestion reaction volume was 20 µL, and the reaction was performed at 37 °C for 15 min.

The success of digestion was confirmed by running the digested samples (20 µL each) on 1% agarose gel. The restriction fragment that contains ToTK1ΔC25-His has a size of ~800 bp, whereas the larger restriction fragment that contains the linearized pET-24a(+) vector has a size of ~5 kb. Both restricted fragments were excised from the gel on UV table, and subsequently isolated and purified using GeneJET Gel Extraction Kit according to the protocol provided. About 15 ng of the ToTK1ΔC25-His fragments was ligated with ~75 ng of the linearized vector based on the molar ratio over the vector 5:1 for each 1 Unit of T4 DNA ligase. The ligation reaction was performed at room temperature overnight. The reaction was terminated by enzyme inactivation by heating at 65°C for 20 min.

About 10 µL of the ligation mix was used to transform 50 µL of TOP10 *E. coli* cells according to the instructions provided. The transformants were selected on LB-agar plates containing kanamycin. The resulting expression plasmids contain the C-terminal His-tagged ToTK1ΔC25 were inoculated in LB media containing Kanamycin and subsequently isolated and purified as described previously.

Expression plasmid for the N-terminal His-tagged ToTK1ΔC25

Unlike the previously constructed C-terminal construct, the N-terminal His-tagged ToTK1ΔC25 construct was generated by PCR using three different overlapping forward primers (P1F, P2F, P3F) containing (in the 5'–3' direction) *Bam*HI and *Nde*I restriction sites, a 8×His tag, and a tobacco etch virus protease (TEV) site, as well as one reverse primer (P2R) containing a *NotI* restriction site. Refer to Appendix I for the full sequences of the primers used.

The toTK1ΔC25 cDNA, obtained in a pGEX-2T plasmid, was used as a DNA template for the amplification reaction. The final volume of the PCR reaction mix was 50 µL containing the following components: 1 µL DNA template, 10 µL of each primer (10 µM), 10 mM dNTPs, 0.5 µL of *Phusion* DNA Polymerase and 10 µL of 5 x *Phusion* buffer. The PCR program was set as follows: 98 °C for 30 s, followed by 35 cycles of 98 °C (10 s), 56 °C (30 s), and 72 °C (90 s), and 72 °C for 10 min. The amplification result was analyzed by running the samples on 1% agarose gel, and the fragment size was ~750 bp. The PCR product was purified using a GeneJET PCR Clean-up Kit according the protocol provided.

The purified PCR product was digested with FastDigest™ enzymes *Nde I* and *Not I* and then introduced into a pET-24a(+) vector using T4 DNA ligase as described before. The ligation mix was used to transform 50 µL of TOP10® *E. coli* cells and the transformants were selected on LB-agar plates supplemented with kanamycin. The resulting expression plasmids carry the N-terminal His-tagged ToTK1ΔC25 were amplified, isolated and purified as described before.

The success of cloning was confirmed by digestion test of each expression vector (carrying the C- and N- terminal construct) using the restriction enzymes *Nde I* and *Not I.* The confirmed expression clones were further analyzed by Automated Sanger DNA sequencing (Eurofins Genomics) in order to confirm the presence of the complete target sequence (ORF) of ToTK1ΔC25 gene with the added tags as expected.

Transformation of electro-competent *E. coli* strains

Three different *E. coli* strains have been used for expression of ToTK1ΔC25. Those expression hosts are: Clear Coli®, Rosetta™ 2 and Tuner™ (DE3). About 50 µL cell suspension of each *E. coli* strain were mixed with ~100 ng of the expression plasmids carrying the ToTK1ΔC25 constructs (N- and C-terminal constructs). The mixture was carefully transferred into electroporation cuvettes. The transformation of the *E. coli* cells was performed by means of electroporation using Bio-Rad MicroPulser Electroporation

Apparatus with the following parameters: 2.5 kV, 25 μF and 200 Ω. The transformants were selected on LB-agar plates containing kanamycin.

Expression of recombinant ToTK1ΔC25 in *E. coli*

Expression trials

Several expression trials were performed according to LP3 protocol in order to optimize and determine the best cultivation conditions for ToTK1ΔC25 production in respect of the type of *E. coli* host (Clear Coli®, Rosetta™ 2 and Tuner™ DE3), cultivation/expression temperatures (30°C and 25°C), inducer concentration (0.4 and 0.8 mM IPTG), and the type of cultivation media (LB and TB media). Individual *E. coli* expression clones that carry expression plasmids were streaked out on LB agar plates, supplemented with the appropriate antibiotics based on each *E. coli* strain, and incubated at 37°C overnight. The growing colonies on the plates were re-suspended in 2-3 mL LB media, and the OD_{600} of the cell suspension was measured after a 200-fold dilution with the same growth media.

Twenty-five mL of growth medium supplemented with the appropriate antibiotics in 150-mL baffled Erlenmeyer flasks were inoculated with the cell suspension to an initial OD_{600} of 0.1-0.2. Cultures were incubated at 37 °C and 200 rpm for growth and the OD_{600} was monitored until it reached a value between 0.6 and 1.0 units in case of LB media or between 1.0 and 2.0 units in case of TB media. The protein expression was induced with the respective IPTG concentration and allowed to continue overnight at the desired expression temperature with shaking at 200 rpm. One-mL samples of the each culture were collected at appropriate time points throughout the growth phases and subsequently centrifuged at 10 000 rpm for 5 min at 4 °C. The pellets were re-suspended in 250 μL of lysis buffer and sonicated at 20 % amplitude for 3 × 10s using pulses of 3s. The cell lysate was clarified by centrifugation at 15 000 rpm for 45 min at 4 °C and the supernatants containing soluble protein were collected and stored at -20 °C until analysis.

Large scale expression

A. **Protein Expression in 5-liter flasks:** a total volume of 6 Liters of growth medium were prepared (2 L in each 5L Erlenmeyer baffled flask) for production of recombinant ToTK1ΔC25 under the optimized conditions for best protein expression. Those optimal conditions that resulted from the expression trials are: *E. coli* Rosetta™ 2 as expression host, C-terminal his-tagged construct, cultivation temperature at 25°C, TB media, shaking speed at 150 rpm and 0.4 mM of IPTG. Cells were first grown at 30°C until the desired OD_{600} for induction was reached (1.0-2.0 units). After induction, the temperature was set at 25°C and the protein expression was allowed to continue overnight. After ~18 hours of induction, cells were harvested by centrifugation at 6000 × *g* for 15 min at 4°C. Cell pellets were collected, weighed and stored at -80°C. Cell disruption was performed by re-suspending the pellets in 25-30 mL lysis buffer supplemented with protease inhibitor cocktail.

The cell re-suspension was passed through a French press cell two times to ensure full cell disruption. The cell lysate was clarified by centrifugation at 13000 × *g* for 30 min at 4°C. The supernatant that contains the soluble protein was collected and passed through a 0.45 µm Whatman cellulose acetate filter to remove all other remaining cell debris.

B. **Protein Expression in one-liter flasks:** a total volume of 1 liter of growth medium was prepared and distributed into 4 Erlenmeyer baffled flasks, each flask contains 250 mL of media. In this experiment, the production of recombinant ToTK1ΔC25 was performed under slightly modified expression conditions, which are: Rosetta™ 2, C-terminal construct, 25°C, 180 rpm, LB or TB media, and 0.2 or 0.4 mM IPTG. The cells were grown at 30°C until the OD_{600} reached the desired value for induction. Then, the protein expression was allowed to take place overnight and lately the cells were harvested by centrifugation as described before. The cell biomass was weighed and stored at -80°C. The following steps for cell lysis and disruption were performed as described previously.

Purification of recombinant toTK1ΔC25 from *E. coli*

Purification of recombinant toTK1ΔC25 was performed by means of Immobilized Metal ion Affinity Chromatography (IMAC) using 1 mL HisTrap HP column charged with Ni^{+2} ions, and ÄKTA Explorer 100 chromatography system. However, this system had unforeseen malfunctions in the pumps so it was replaced by a simple peristaltic pump.

The chromatography process was performed in two phases within 2 days. The first phase of the purification process was performed by applying the clarified and filtered cell lysate (crude extract CE) into HisTrap HP column filled with sepharose and pre-equilibrated with 20 mL binding buffer. The Flow Through fraction containing unbound particles was collected and referred as FT-1. In order to remove other contaminants that bound to the column and also to stabilize the bound enzyme, two ATP washing steps were performed by recirculating binding buffer supplemented with $MgCl_2$ and ATP with a final concentration of 10 mM each. The fractions that were resulted from each washing step were collected separately and called ATP-W1 and ATP-W2. The His-tagged toTK1ΔC25 was eluted from the column by adding elution buffer (EB) containing 300 mM Imidazole. Instant dilution of the eluted fraction was performed by collecting the eluate drops as they abandoned the column in 180 mL cold dilution buffer (DB) supplemented with 1 mM DTT, so the collected eluted fraction was 10-fold diluted. A washing step with a higher concentration of imidazole (500 mM) was introduced to ensure that no proteins remained bound to the column.

In order to cleave the histidine tag from the eluted enzyme, 1 mg of tobacco etch virus (TEV) protease was added for every 40 mg of eluted ToTK1ΔC25-His. The cleavage reaction (CR) was allowed to take place overnight at 16°C. The next day, a second chromatography process was performed by applying the cleavage reaction content into same pre-equilibrated HisTrap HP column. The cleaved toTK1ΔC25 passed the column without

binding to the matrix. Therefore, the flow-through containing the cleaved protein was collected. Afterwards, the column was washed with a high concentration of imidazole (500 mM) to remove the uncleaved toTK1ΔC25 and the His-tagged TEV protease, which are bound strongly to the column due to the histidine tag.

Recombinant Production of ToTK1ΔC25 by Baculovirus Expression Vector System (BEVS)

Generation of recombinant Baculoviruses

I. **Construction of donor plasmids:** recombinant donor plasmids had been constructed previously (by the supervisor L.S.C) by subcloning either C- or N-terminal His-tagged ToTK1ΔC25 into the the *BamHI/NotI* site of pFastBac™ Dual GFP, which also codes for enhanced green fluorescent protein (eGFP).

II. **Generation of recombinant bacmids:** the Bac-to-Bac® Baculovirus Expression System by Life Technologies provides a rapid and efficient method to generate recombinant baculoviruses (Ciccarone et al., 1997). The first step was to generate recombinant bacmids using *E. coli* DH10Bac™ strain as a host for cloning. About 100 ng of the donor plasmid were used to transform 20 µL of *E. coli* DH10Bac™ by means of heat-shock method according to the instructions provided. The colonies containing recombinant bacmids were selected based on white/blue selection method by plating on LB-agar containing three types of antibiotics (tetracycline, kanamycin and gentamycin), the chromogenic substrate Bluo-gal (Invitrogen) and the inducer IPTG in appropriate concentrations as recommended. The colonies bearing the recombinant bacmids appeared white, whereas the ones that didn't contain the bacmids appeared blue. Two consecutive selection rounds of the white colonies were performed by plating those colonies on LB selection plates followed by incubation overnight at 37°C.

III. **Bacmid purification:** single white colonies that contain the recombinant bacmids were used to inoculate aliquots of 5 mL LB media supplemented with all three antibiotics, and allowed to grow at 37°C overnight. Plasmids were isolated and purified using PureLink® HiPure Plasmid Miniprep Kit according to the protocol provided.

IV. **Analysis of recombinant bacmids by PCR:** the aim of this analysis was to verify the presence of the target gene in the recombinant bacmid. The PCR reaction mix (20 µL) was composed of the following components: 0.4 µL of the purified bacmid as a template, 0.1 µL of both forward and reverse pUC/M13 sequencing primers (0.5 µM each), 0.2 µL DreamTaq DNA Polymerase, 2 µL of 5× DreamTaq Green Buffer, 0.4 µL of 10 mM dNTPs, and 16.8 µL ddH_2O.

The PCR program was set as follows: 95 °C for 3 min, followed by 35 cycles of 95 °C (45 s), 55 °C (45 s), and 72 °C (5 min), and finally 72 °C for 7 min. The PCR product was analyzed on 0.8% agarose gel. The expected size of the bacmid

alone is 300 bp. A bacmid transposed with empty pFastBac™ Dual is 2560 bp plus insert size (~675bp).

V. **Transfection of insect cells:** the purified and verified recombinant bacmids were used to transfect Sf9 insect cells *(Spodoptera frugiperda)* growing in EX-CELL® 420 serum-free medium to generate recombinant baculovirus. To facilitate the transfection process, FlashFECTIN reagent was used as instructed. Viable Sf9 cells were diluted to a final concentration of 0.5×10^6 cells/mL, and seeded into 6-well plates (2 mL culture/well). Cells were left to adhere to the surface of the well for ~1 hour until they form a ~30-50% confluent. The bacmids were heated in a heat-block at 60°C for 10 minutes before first usage. In order to generate recombinant baculoviruses, a mixture of 100 µL EX-CELL medium and 20 µL of the corresponding bacmid was combined with a mixture of 100 µL EX-CELL medium and 9 µL FlashFECTIN. The combined transfection mixture was incubated at room temperature for 15 minutes, and subsequently added drop-wise to the 6-well plate. In each plate, two controls were used. One well for the flashFECTIN control that didn't have bacmids, and the other well for the media control which had only insect cells and media. The plates were put inside plastic bags together with damp towels (to prevent drying) and incubated at 27 °C for 7 days. The transfection process was monitored by detecting GFP expression using an inverted fluorescence microscope. At the end of the incubation period, the recombinant baculovirus was recovered by centrifugation at 1000 × g for 10 min. The supernatant containing the recombinant baculovirus is referred as P1 supernatant.

VI. **Baculovirus amplification:** this step was performed in order to generate high-titer baculoviral stock. Thus, 20-mL cultures of Sf9 cells (2×10^6 cells/mL) in EX-CELL medium were infected with 500 µL P1 supernatant and incubated for 1 week at 27 °C and 100 rpm. After 4 days of incubation, 500 µL samples were taken from the cultures to screen for thymidine kinase activity in the crude extract. At the end of the incubation period, cell viability was checked to confirm cell death by trypan blue staining and counting. Insect cell debris were removed by centrifugation (1000 × g, 10 min) and the baculoviral supernatants (P2) were collected in sterile tubes and stored in the dark at 4 °C.

Baculovirus quantification by Endpoint dilution assay

This assay was performed in order to determine the virus titer in terms of infectious viruses per mL (pfu/mL) in the P2 supernatants and subsequently to calculate the multiplicity of infection (MOI). Serial dilutions of P2 supernatants in EX-CELL media ranging from 10^{-2} to 10^{-9} were prepared with a final volume of 99 µL. Viable Sf9 cells in EX-CELL medium were diluted to a density of 0.2×10^6 cells/mL and 890 µL of diluted cells were added to each of the tubes containing virus dilutions. The baculovirus/insect cell mixtures were seeded into a 96-well plate (100 µL/well, 9 replicates for each dilution). As positive controls, two wells were seeded with Sf9 cells mixed with undiluted virus.

As negative controls, two wells were seeded with insect cells and growth media only. The plates were incubated for 7 days at 27 °C inside a plastic bag containing a damp paper towel. After one-week incubation, the GFP expression was screened by fluorescence microscopy in each of the wells. The number of positive (GFP expression) and negative wells (no GFP expression) was counted for each dilution of the virus and the titer (pfu/mL) was calculated as described by King and Possee (King & Possee 1992).

Expression of recombinant ToTK1ΔC25 in insect cells

Expression test was performed in two insect cell lines: Sf9 (cultivated in SFX medium) and High Five™ (cultivated in SFX medium). Both insect cell cultures were diluted to a final density of 1.5 × 10^6 cells/mL, and infected with the proper volumes of **P2** supernatants to give multiplicities of infection (MOI) of 0.5 and 5 pfu/insect cell in a final culture volume of 5 mL.

As a background control for kinase activity, an empty baculovirus which expresses eGFP only was used along with the constructs to be expressed. This control is called (MOCK control). Protein expression was performed in 50 mL TPP bioreactors under the conditions: 25 °C and 200 rpm. After 72 h of cultivation, the cells were harvested by centrifugation at 4500 × g for 20 min (4 °C) and the pellets were stored at -20 °C until analysis. To release the intracellular protein, 2 mL of insect cell lysis buffer supplemented with 1 mM DTT and 1X protease inhibitor cocktail was added to each pellet and the cell suspension was sonicated on ice (20% amplitude, 3 cycles of 10 pulses each, 3 seconds per pulse). Cell lysates were clarified by centrifuging at 16 000 × g for 10 min at 4 °C and the supernatants were stored at -20 °C until further analysis of the expressed protein.

Protein analysis

Bradford assay

The protein concentrations of the different protein samples were determined by Bradford assay (Bradford, 1976) based on Bovine Serum Albumin (BSA) as a standard.

SDS-PAGE analysis

The size of the target protein was analyzed by SDS-PAGE. First, 40 µL of protein samples were mixed with 10 µL of 5× SDS loading buffer. However, samples that have high protein concentration were diluted with water and mixed with the loading buffer accordingly. Then, samples were incubated in a heat-block at 95°C for 10 min. The tubes were cooled down on ice for a few minutes and subsequently, ~10 µL of each sample was loaded on 12% polyacrylamide Mini-PROTEAN® TGX™ precast gels and run for 10 min at 100 V and 40 min at 180 V. As a marker for molecular weight determination, 5 µL of prestained protein ladder PageRuler™ (10-180 kDa) was loaded in the gel. Linearized proteins were separated based on their size and lately visualized by staining with Bio-Safe Coomasie according the manufacturer's instruction. The gels were scanned and analyzed by a Bio-Rad ChemiDoc MP apparatus using the Image Lab software.

Western Blotting

Protein samples were also analyzed by Western blotting followed by immunodetection in order to detect the His-tagged recombinant ToTK1ΔC25 using antibodies raised against the histidine tag. Eighty microliters of clarified cell lysates were mixed with 20 µL 5× loading buffer. A 1:10 dilution of His-tagged TEV protease was used as a positive control for the screening. Samples were run on a 12% polyacrylamide Mini-PROTEAN® precast gel for 10 min at 100 V and 35 min at 200 V. Protein transfer was performed for 7 min in a Trans-Blot® Turbo™ Transfer System using a PVDF membrane pre-soaked in 96% ethanol and a pair of cotton fiber papers pre-soaked in Transfer Buffer according to the manufacturer's instructions. The blotting process was performed in several steps. First, the blocking step was performed by incubating the membrane with PBS containing 6% milk for ~1 hour at room temperature, followed by washing with PBS for 5 min. Then, the membrane was incubated with a primary antibody solution (Mouse Anti-His6 1:3000 dilution in PBS) for 1-2 h at room temperature followed by washing with PBS three times (3 × 5 min). Afterwards, the membrane was incubated with a secondary antibody solution (Goat Anti-Mouse IgG-Peroxidase 1:20 000 dilution in PBS containing 3% milk) at 4°C overnight. The following day, the membrane was washed with PBS three times as before and subsequently incubated with 3 mL of a 1:1 v/v mix of Luminol enhancer solution and peroxide buffer for a short time (5-10 min). Chemiluminescence was detected with a Biorad ChemiDoc MP station and several pictures were taken at certain time points and further analyzed by Image Lab software.

Thymidine kinase activity assay

The activities of thymidine kinase were measured using radioactive labelled substrate. The time course of the reaction was followed by transferring aliquots of the reaction mixture at four time points to anion-exchange paper disks (DE-81) which are made of DEAE (diethylaminoethyl) cellulose.

First, the substrate was prepared, according to the provided protocol, by diluting radio-labelled ^{3}H-methyl-thymidine (20 000 mCi/mmol) with unlabeled thymidine and deionized water to produce a thymidine stock with a final concentration of 5 mM. Different samples of the recombinant ToTK1ΔC25 were diluted (when needed) in enzyme dilution buffer composed of 50 mM Tris-HCl, pH 7.5, 1 mM CHAPS, 3 mg/mL bovine serum albumin (BSA), 5 mM dithiothreitol (DTT) and 5 mM $MgCl_2$.

The assay mix was composed of the following reagents: 50 mM Tris-HCl, pH 7.5, 10 mM dithiothreitol (DTT), 3 mg/mL bovine serum albumin (BSA), 2.5 mM $MgCl_2$, 2.5 mM ATP, 0.5 mM CHAPS, and 100 µM tritium-labeled thymidine as the substrate. Each assay had a total volume of 50 µL (40 µL assay mix + 10 µL enzyme dilution).

First, 10 µL of each enzyme dilution was added into a glass tube that was put on ice. Then, the reaction was initiated by adding 40 µL of the assay mix into the appropriate assay tubes, vortexing and placing them at 37°C water bath at a certain time (considered as a starting point). After 4, 8, 12 and 16 min, 10 µL from each assay tube was taken and put on the previously-labeled DE-81 filter paper disks accordingly. Finally, 5 µL samples were taken from 4 random tubes and put on the paper disks. These are the standards to use for calculations to convert radioactivity measurements (cpm) into concentration (pmol).

The paper disks were washed (except the standards) three times each for 5 min with 5 mM ammonium formate to remove the non-phosphorylated substrate. Additional washing step was performed for 5 min with only deionized water followed by sorting the papers and drying them on a paper towel. Individual filter papers were placed inside plastic liquid scintillation vials (one for each paper) and the radioactive products were extracted by addition of 0.75 mL of elution solution (0.2 M KCl, 0.1 M HCl) and shaking for ~30 min. Finally, 4.5 mL of scintillation fluid was added to each vial and the mixture was mixed by inversion and put in scintillation counter. Radioactivity was determined by measuring radiations counts per minute (cpm) for each vial by applying the appropriate program.

By making a linear plot of cpm versus time, the slope Δ1 can be estimated and the enzyme activity can be determined based on the following formula:

Δ1× ((total assay volume/ (vol. of enzyme × vol. added on filter) **×** (pmol of dThd/ average cpm)**)**

<u>One unit (u) of kinase activity</u> was defined as 1 nmol of thymidine-5'-monophosphate (dTMP) formed per minute. One Unit (U) of activity equals 1000 units and corresponds to 1 µmol of product formed per min.

Results

Generation of expression plasmids for the mutant ToTK1

Both N- and C-terminal construct have been successfully cloned and purified from *E. coli.* The correctness of the inserts was confirmed by Automated Sanger Sequencing and restriction analysis. The confirmed clones that proved to have the correct open reading frames (ORF) were selected to transform *E. coli* expression hosts (Fig. 5)

The C-terminal His-tagged construct was successfully generated from the clone containing the recombinant plasmid pET-24a(+) bearing the insert ToTK1ΔC25-His, with a DNA concentration of 58 ng/μL and A_{260}/A_{280} = 2.09. The sequence analysis of the C-terminal construct showed that the ORF was consisted of 678 bp which encodes a protein of 226 amino acids, which represent the size of ToTK1ΔC25 plus a 7-amino acid TEV cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Ser) followed by 8×His tag and three additional spacer residues to protect the C-terminal tag. The molecular weight of the fusion protein is 25.3 kDa.

On the other hand, the N-terminal His-tagged construct was successfully generated from the clone containing the recombinant plasmid pET-24a(+) bearing the insert His-ToTK1ΔC25, with a DNA concentration of 50 ng/μL and A_{260}/A_{280} = 1.90. The sequence analysis of the N-terminal construct showed that the ORF was consisted of 699 bp which encodes a protein of 233 amino acids, which consists of 8×His tag followed by 8 amino acid residues as spacer followed by 7 amino acids for TEV cleavage site and the ORF of ToTK1ΔC25. The molecular weight of the fusion protein is 25.9 kDa.

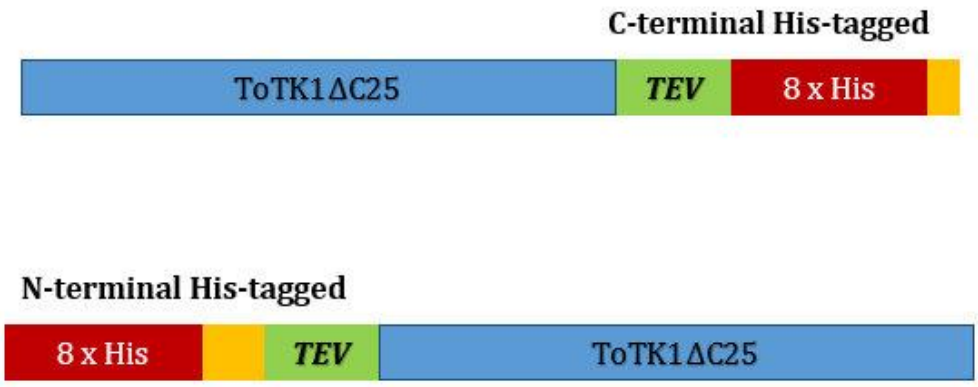

Figure 5 Graphical representation of the C- and N-terminal His-tagged ToTK1 constructs. The complete fusion protein of each construct has the following domains: ToTK1ΔC25, C-terminal truncated tomato thymidine kinase 1 open reading frame. TEV, tobacco etch virus protease cleavage site. 8xHis, oligohistidine tag. Orange bars, additional amino acids as spacer / C-terminal protection.

Expression of recombinant ToTK1ΔC25 in *E. coli*

Expression trials

All the three *E. coli* strains (Rosetta™ 2, Clear Coli® and Tuner™ DE3) were successfully transformed with both C- and N- terminal constructs and several expression trials were performed to screen for the best performing strain and expression conditions. SDS-PAGE analysis showed that no expression was observed in all of the three *E. coli* strains transformed with the N-terminal Construct. However, among the *E. coli* strains transformed with the C-terminal construct, Rosetta™ 2 showed the best performance with intense bands when compared to Tuner™ DE3 which showed less intense bands (Fig. 6A and 6B). On the contrary, Clear Coli® had shown no expression compared to Rosetta™ 2 and Tuner™ DE3 (data not shown). For both Rosetta™ 2 and Tuner™, expression was performed at 25 °C with 0.4 mM IPTG resulted in target protein bands ~18-21 h post-induction. Target protein bands were not detected before induction, suggesting that the promoter is not leaky. Expression at 30 °C resulted in very less intense bands and lower OD_{600} compared to the ones expressed at 25°C in all cases.

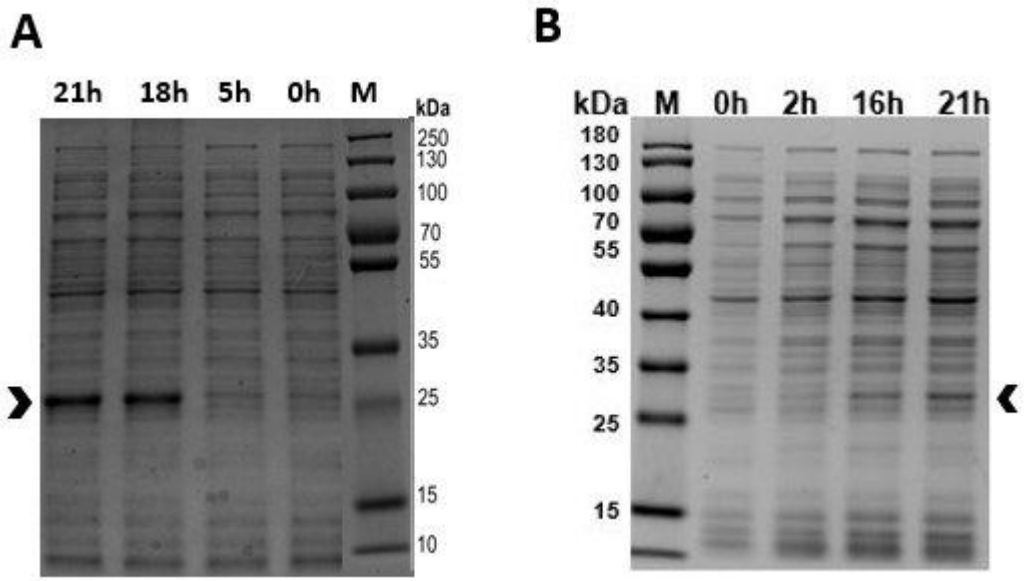

Figure 6 SDS-PAGE analysis of recombinant ToTK1ΔC25 expression trials in *E. coli*. (A) C-terminal His-tagged ToTK1ΔC25 (25.3 kDa) expressed in *E. coli* Rosetta 2. (B) C-terminal His-tagged ToTK1ΔC25 expressed in *E. coli* Tuner DE3. In both cases, expression was induced at t = 0 h with 0.4 mM IPTG at 25 °C in LB medium. Proteins were visualized by Coomassie blue staining. The approximate sizes of the molecular weight markers (lane M) are indicated.

After screening for the best performing *E. coli* strain, which was Rosetta™ 2 bearing the C-terminal construct and expressing at 25°C, another expression trial was performed for the chosen strain (Rosetta™ 2) in order to determine if the recombinant production of ToTK1ΔC25-His could be further improved by changing IPTG concentration (0.4 mM or 0.8 mM IPTG) and the type of growth media (LB or TB) (**Fig. 7**). As a result, modifying the IPTG concentration had no noticeable effect on the target protein level.

However, cultivation in TB medium had a large effect on the final OD_{600} values obtained (higher bacterial biomass). The final OD_{600} for the cultures growing in LB medium after ~16 hours of induction was 2.4 units, whereas in TB medium the OD_{600} reached 8.9.

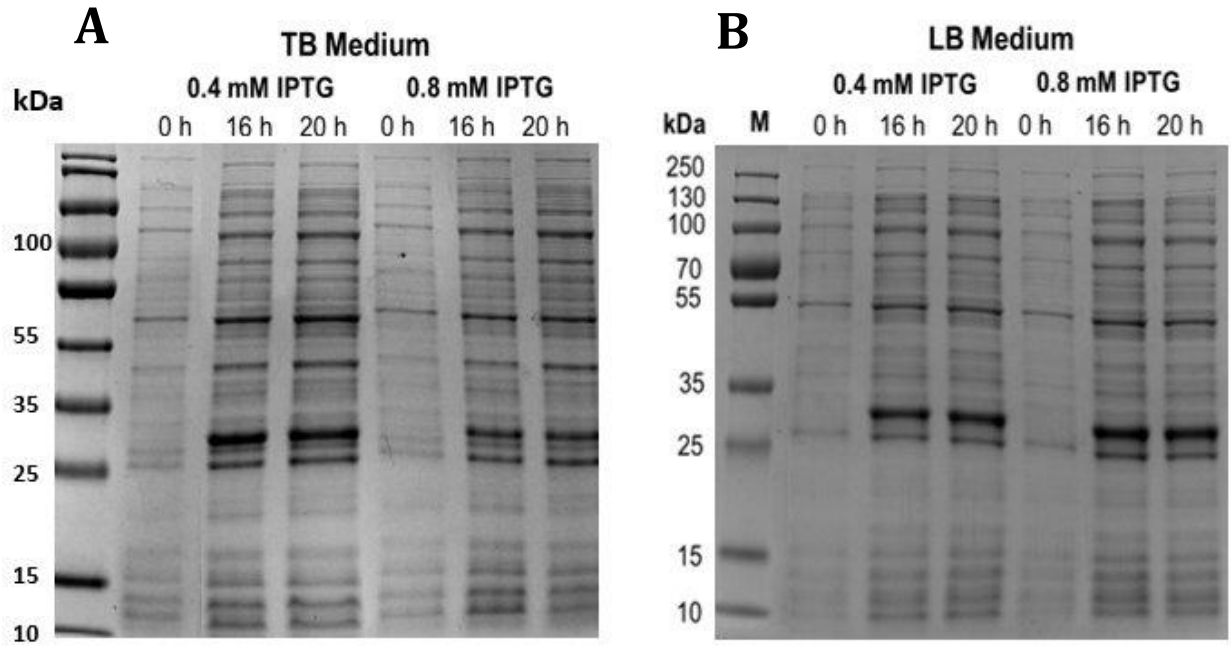

Figure 7 **SDS-PAGE analysis of recombinant ToTK1ΔC25 expression trials in *E. coli Rosetta 2*.** (A) C-terminal His-tagged ToTK1 (25.3 kDa) expressed in *E. coli* Rosetta 2 cultivated in TB Medium. (B) C-terminal His-tagged ToTK1ΔC25 expressed in Rosetta 2 cultivated in LB medium. In both cases, expression was induced with either 0.4 or 0.8 mM IPTG (at t = 0 h) and cells were cultivated at 25 °C. Proteins were visualized by Coomassie blue staining. The approximate sizes of the molecular weight markers (lane M) are indicated.

Based on the results from the expression trials performed, the following protein expression conditions were selected for scale up expression: Rosetta™ 2 bearing the C-terminal His-tagged ToTK1ΔC25 construct and induction with 0.4 mM IPTG at 25 °C in TB medium.

Large-scale expression of ToTK1ΔC25-His in *E. coli* Rosetta™ 2

The protein expression of ToTK1ΔC25-His was scaled up based on the optimal conditions obtained from the expression trials. Here, the protein was successfully expressed in 5-liter Erlenmeyer flasks containing 2 liters of TB medium. Cells were harvested by centrifugation after 18 h from induction. The final average OD_{600} obtained was 3.38 ± 0.28 and the average cell wet weight measured was 4.53 ± 0.41 g per liter of culture. By comparing the OD_{600} values with the values recorded in the expression trials, we conclude that the scaling up from 20 ml of culture media to 2 L didn't result in higher OD_{600} values as expected. So, another smaller scale protein expression was performed in 1-liter flasks containing 250 mL of growth media in order to improve aeration and also to increase the biomass yield. These smaller scale cultures reached a final OD_{600} of 4.6 units for LB cultures with an average cell wet weight of 6.4 ± 0.15 g per L culture, and 8.9 units for cells growing in TB medium with an average cell wet weight of 9.2 ± 0.18 g per L culture. Varying the concentration of IPTG had no effect on biomass.

Purification of ToTK1ΔC25-His from *E. coli* Rosetta™ 2

Successful preliminary purification of recombinant ToTK1ΔC25-His from the crude extract was performed by means of IMAC method using Ni^{+2} affinity column. The **figure 8** shows the SDS-PAGE analysis of the purification procedure. An intense band that represents the target protein with a molecular weight of ~26 kDa can be observed in the crude extract (CE), but not in the flow-through (FT) fraction and it becomes significantly improved in the eluted fraction (E). However, the SDS-PAGE photo shows some protein degradation and contaminants (70-100 kDa) in the eluted fraction. Therefore, two rounds of ATP washing procedure with ATP/$MgCl_2$ was introduced in attempt to remove the non-target contaminating proteins (70-100 kDa in size) that are co-purified with ToTK1ΔC25-His, which are likely to correspond to the chaperones DnaK and members of the Clp family of *E. coli* (Carroni 2014).

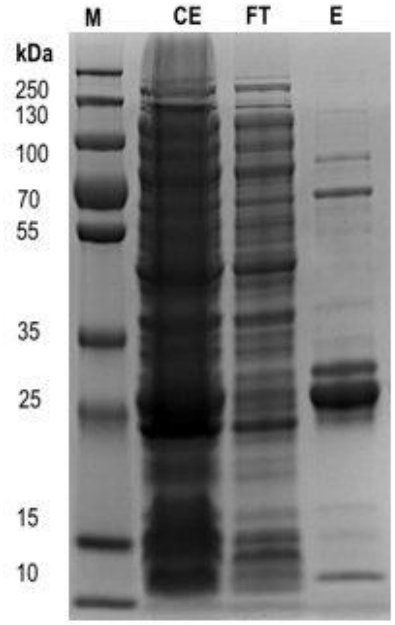

Figure 8 SDS-PAGE analysis of recombinant ToTK1ΔC25 preliminary purification from *E. coli* Rosetta 2 cells cultivated in 2 L TB medium. The approximate sizes of the molecular weight markers (lane M) are indicated. CE, crude extract. FT, crude extract flow-through after Ni^{2+} affinity column. E, pooled ToTK1ΔC25 fractions eluted from the Ni^{2+} column.

Although recombinant ToTK1ΔC25-His was successfully expressed and purified. The purified protein has a high tendency to form aggregates and precipitate after a short time from elution. However, several strategies were applied in an attempt to improve the solubility of the recombinant ToTK1ΔC25-His, including using desalting PD-10 columns and Vivaspin™ centrifugal concentrators in order to exchange the imidazole in elution buffer for binding buffer which has lower concentration of Imidazole (20 mM), or for Tris buffer supplemented with 100 mM NaCl.

Furthermore, additives such as detergents (CHAPS), enzymatic cofactors (ATP/Mg) and reducing agents (DTT) were tested separately and in various combinations to determine if they could improve the solubility of ToTK1ΔC25.

Buffer exchange delayed precipitation for several hours, but did not prevent it. Moreover, the buffer exchange process was very slow and inefficient because of blockage in the packed bed (filter) of the PD-10 column caused by protein precipitates. Dialysis was suggested but not attempted as it is time consuming process. The additives (CHAPS, DTT and ATP/$MgCl_2$) did not improve solubility. It just delayed and partially decreased precipitation which occurred after ~ 2 days.

The chromatogram (not shown) generated from previous purification trial using ÄKTA Explorer 100 purification system, revealed that the ToTK1 peak elutes from the Ni^{2+} affinity column when the imidazole concentration is ~250 mM which corresponds to ~50% of the original Elution Buffer. Therefore, in order to minimize protein precipitation, the concentration of imidazole in the elution buffer was reduced from 500 mM to 300 mM, and the eluted protein was instantly diluted ten-fold in Dilution Buffer as it left the column. As a result, the final imidazole and NaCl concentrations was decreased to 30 mM and 130 mM, respectively. Moreover, the oligohistidine tag was cleaved by incubating with TEV protease and by reapplying the cleavage reaction mix into the IMAC column.

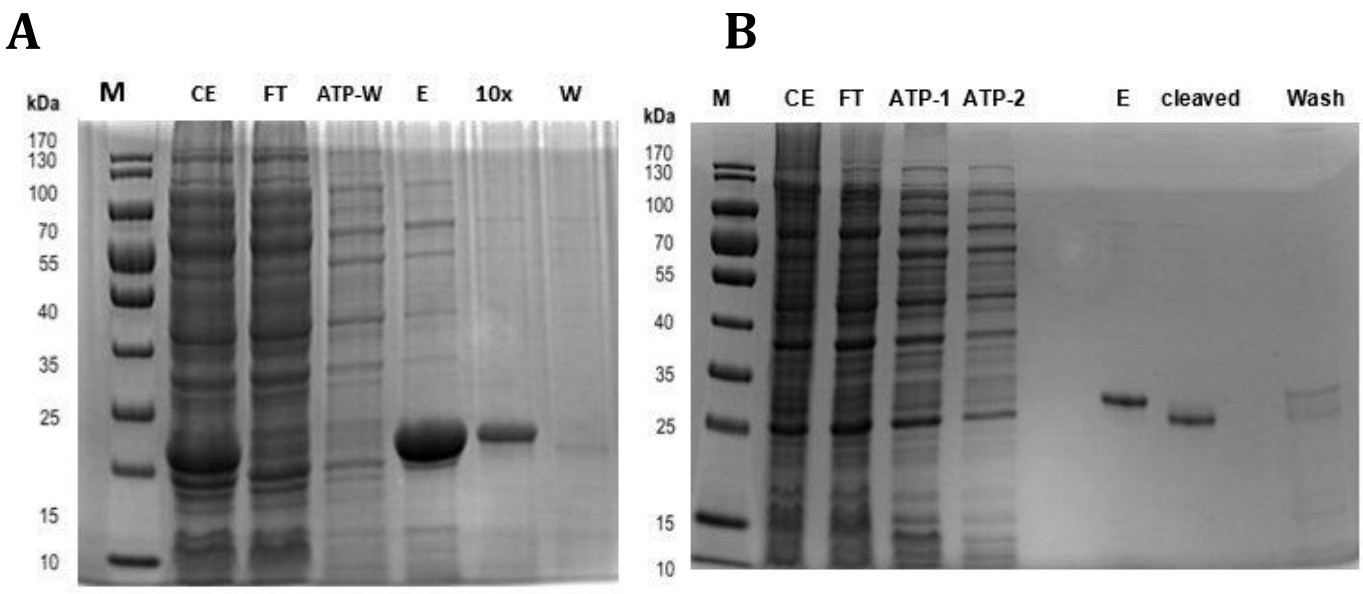

Figure 9 **SDS-PAGE analysis of recombinant ToTK1ΔC25 purification from *E. coli* Rosetta 2.**
(A) Purification of ToTK1ΔC25 from Rosetta 2 cultivated in 250 mL of TB medium. CE, crude extract. FT, flow through. ATP-W, eluate from ATP/$MgCl_2$ washing step. E, eluted fraction undiluted. 10x, eluted fraction diluted 10 folds. W, washing with imidazole.
(B) Purification of ToTK1ΔC25 from Rosetta 2 cultivated in 250 mL of LB medium. CE, crude extract, FT, flow through. ATP-1 and APT-2, eluates from ATP/$MgCl_2$ washing steps. E, ToTK1ΔC25 fraction eluted from the Ni^{2+} column and diluted 10 folds. Cleaved, fraction after TEV cleavage with TEV protease. Wash, wash with imidazole. Approximate sizes of the molecular weight markers (lane M) are indicated.

The SDS-PAGE analysis shows that the purification process had been improved significantly after washing with ATP/$MgCl_2$ and removing the oligohistidine tag by TEV protease. Figure (9) shows the SDS-PAGE analysis of the improved purification process of recombinant ToTK1ΔC25-His produced by *E. coli* Rosetta™ 2 in TB medium (Fig. 9A) and LB media (Fig. 9B). These results have shown that the contaminating proteins have been reduced in the eluted fraction, and the ATP washing steps took away those contaminants effectively. The SDS-PAGE photo showed clearly improved purity of the target protein in the eluted fraction with estimated homogeneity of ~90%.

The following table (**Table 2**) shows the protein concentrations measured by Bradford Assay. The concentrations and yields of proteins produced from 250 mL of TB medium were higher than those produced in LB medium. This fact demonstrates that, although higher biomass and protein yields can be achieved if richer media is used, the proportion of target protein in the total protein fraction increases if less nutritious media is employed. A significant portion of the target protein is lost due to precipitation during TEV cleavage reaction which took place over night at room temperature (~16 °C).

Table 2 Protein concentrations and yields resulted from 250 mL of culture medium

Fraction	Cultivation Medium	Protein concentration (mg/mL)	Protein amount (mg)	Yield (mg/L)
Raw Extract (RE)	TB	1.25	312.5	1250
Purified uncleaved	TB	0.2	50	200
Purified cleaved	TB	0.09	22.5	90
Raw extract (RE)	LB	0.8	200	800
Purified uncleaved	LB	0.11	27.5	110
Purified cleaved	LB	0.06	15	60

Although ToTK1ΔC25 was produced and purified successfully from *E. coli*, it was not possible to detect significant TK enzymatic activity in ToTK1ΔC25 purified samples, not even in samples that displayed no precipitation. All samples that were measured didn't show accumulation of dTMP over time and no regression pattern observed in the counting data (CPM) (data not shown). However, TK enzymatic activity could only be measured in the crude extract fractions, but it was also present in the samples that contained no ToTK1, such as flow-through samples from the IMAC purification and also the non-induced controls.

Generation of recombinant baculoviruses

Recombinant bacmids bearing the gene coding for ToTK1ΔC25 were successfully created and analyzed by PCR to confirm the presence of the target gene. The PCR products of both C- and N- terminal constructs were analyzed by agarose gel electrophoresis. Three C-terminal His-tagged constructs (named C1, C2 and C3) and one N-terminal constructs (named N3) displayed an approximate size of ~3.5 kb (2650 bp from bacmid DNA plus the coding sequence of ToTK1ΔC25). The recombinant bacmids of the analyzed constructs that proved to have the sequence coding for ToTK1ΔC25 were selected to transfect Sf9 insect cells.

The transfection efficiency of Sf9 cells was followed by its expression of green fluorescent protein (GFP). The figure (10) shows GFP expression in the cells transfected with bacmids bearing C- and N- terminal constructs. On the other hand, there was no GFP expression in the flashFECTIN control (FF Ctrl) or in the control containing growth media only (M Ctrl). Therefore, GFP expression indicates that the transfection process was successful. However, there's no necessary relation between GFP expression and ToTK1ΔC25 expression.

After 3-4 days of transfection with recombinant baculoviruses for viral amplification, it was possible to detect thymidine kinase activities in the crude extracts of Sf9. The following **specific activities** (units/mg total protein) were measured: 154.4 u/mg for C1, 102.5 u/mg for C2, 85.3 u/mg for C3 and 72 u/mg for N3. At the end of the baculovirus amplification process, the **viral titers** calculated in the **P2** supernatants generated from C1, C2 and N3 were as follows: 1.0×10^{7} pfu/mL for the C1 construct, 1.07×10^{8} pfu/mL for the C2 construct, and 3.94×10^{8} for N3.

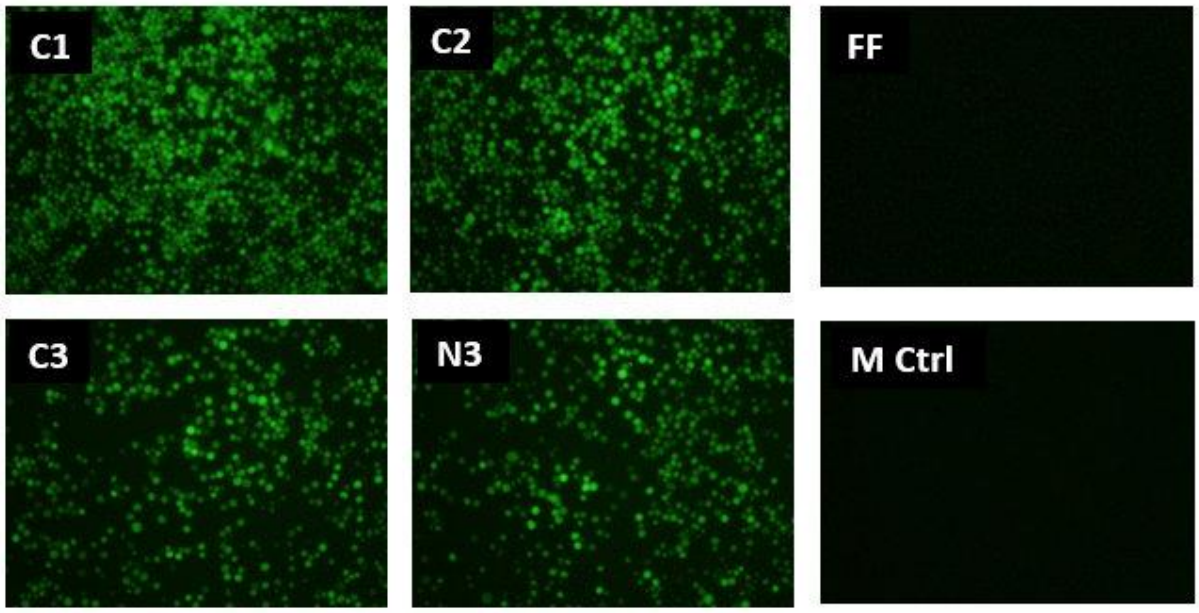

Figure 10 Fluorescence photos due to GFP expression that represents the transfection efficiency of Sf9 insect cells infected with recombinant baculoviruses bearing the ToTK1ΔC25 C-terminal constructs (C1, C2and C3) and N-terminal constructs (N3). Flashfectin (FF) and Negative Controls (Ctrl) are also included.

Expression of recombinant ToTK1ΔC25 in insect cells

Expression test and TK activity

Here, the TK activity for both the wild-type ToTK1 constructs and the truncated (ToTK1ΔC25) constructs were measured together for comparison. The TK activity measurements of the wild-type ToTK1 were performed separately in a parallel project by Juan Manuel Orozco Rodriguez, a fellow student. **Table 3.** Shows the TK activity measurements of insect cell crude extracts with 100 µM dThd as substrate. The TK activity measurements clearly show an increase in detected TK activity compared to the background control (MOCK). Previous testing with undefined MOI had also indicated significant TK activities on day 3 after infection in both Sf9 and High Five™ cell lines (data not shown). For each tested construct and MOI, a higher specific activity (u/mg) was found in Sf9 cells compared to High Five™ cells. Therefore, expression of all constructs in Sf9 cells with an MOI of 5 appears to be favorable. The highest specific activity (896 u/mg) is achieved by infecting Sf9 cells (MOI of 5) with the baculovirus bearing the mutant C-terminal construct (ToTK1ΔC25-C-His).

The total TK activity (units) corresponding to the entire volume of the insect cell culture (5 mL) was also calculated. This value was higher for High Five™ cells than for Sf9 cells, which is not unexpected, since total protein concentration was consistently superior in High Five™ extracts. For both cell lines, an MOI of 5 is more favorable than 0.5. The construct capable of yielding the highest total activity when expressed in High Five™ cells at MOI 5 is ToTK1ΔC25-C-His (2453 units), followed closely by both wild-type constructs expressed under the same conditions (2292 units for N-His-ToTK1 and 2264 units for ToTK1-C-His).

Table 3 TK activity measurements. TK activity was measured in cell extracts from insect cells infected with baculovirus expressing different ToTK1 constructs or eGFP alone (MOCK) 3 d after infection. Two independent expression trials were performed and TK activity was independently determined twice for each trial. The table shows the average of all measurements ± SD normalized to total protein concentration of the crude extracts as specific activities, as well as the total activity, measured per volume (5 mL) of insect cell shaker cultures.

Baculovirus	Cell line	MOI	Specific Activity (u/mg)	Total Activity (u)
MOCK	High Five™	0.5	0.1 ± 0.01	0.3 ± 0.01
MOCK	High Five™	5	0.2 ± 0.01	0.4 ± 0.04
MOCK	Sf9	0.5	0.3 ± 0.08	0.5 ± 0.03
MOCK	Sf9	5	0.1 ± 0.02	0.2 ± 0.04
N-His-ToTK1	High Five™	0.5	47 ± 18	335 ± 4
N-His-ToTK1	High Five™	5	377 ± 23	2292 ± 139
N-His-ToTK1	Sf9	0.5	131 ± 27	319 ± 65
N-His-ToTK1	Sf9	5	523 ± 160	803 ± 246

ToTK1-C-His	High Five™	0.5	33 ± 2	184 ± 9
ToTK1-C-His	High Five™	5	394 ± 63	2264 ± 362
ToTK1-C-His	Sf9	0.5	209 ± 67	454 ± 145
ToTK1-C-His	Sf9	5	459 ± 35	901 ± 69
N-His-ToTK1ΔC25	High Five™	0.5	86 ± 15	572 ± 68
N-His-ToTK1ΔC25	High Five™	5	261 ± 53	1748 ± 236
N-His-ToTK1ΔC25	Sf9	0.5	286 ± 46	613 ± 56
N-His-ToTK1ΔC25	Sf9	5	590 ± 30	1195 ± 77
ToTK1ΔC25-C-His	High Five™	0.5	15 ± 3	89 ± 15
ToTK1ΔC25-C-His	High Five™	5	471 ± 74	2453 ± 167
ToTK1ΔC25-C-His	Sf9	0.5	304 ± 52	633 ± 83
ToTK1ΔC25-C-His	Sf9	5	895 ± 35	1586 ± 113

Protein Analysis by Western Blot

Figure (11) shows the Western blot and His-tag immunodetection analysis to test and compare the levels of protein expression for the respective constructs in Sf9 and High Five cell lines. The constructs C1, C2, C3 and N3 were expressed in Sf9 cells. The same constructs, except C3, were also expressed in High Five cells. The results have shown better expression and higher total protein concentration in High Five extracts compared with Sf9 extracts. However, significant degradation can be observed in the High Five extracts.

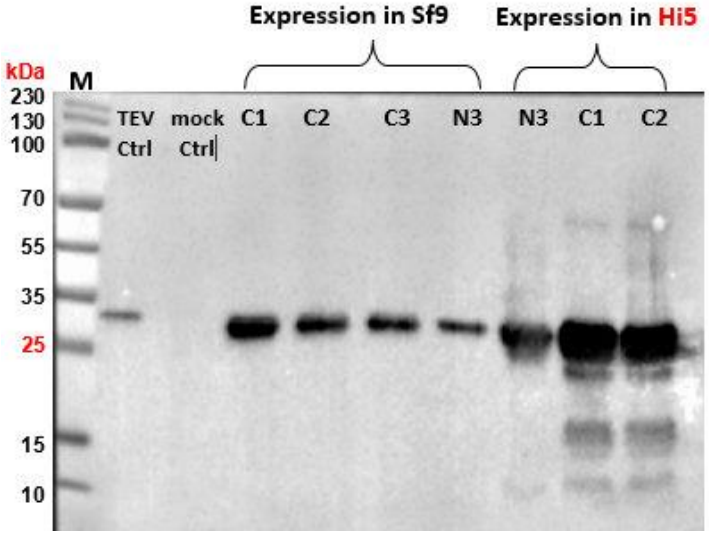

Figure 11 Western blot and immunodetection of ToTK1ΔC25 expression trial in two insect cell lines, Sf9 and Hi5. Purified and His-tagged TEV protease (50 ng) was used as a positive control. As a negative control, an empty baculovirus (eGFP expression only) was used.

Additional Western Blotting analysis followed by immunodetection of the His-tag (**Fig. 12**) was performed on the crude extract samples that displayed the highest specific activity, which were derived from insect cells infected with an MOI of 5 for both wild-type and mutant ToTK1. As can be observed, for both cell lines, the highest expression levels of recombinant ToTK1 correspond to the wild-type C-terminal His-tagged constructs (ToTK1-C-His) and the mutant N-terminal construct (N-His-ToTK1ΔC25) displays the lowest levels.

Interestingly, this pattern of protein expression levels is not reflected in the TK activity measurements. Significant degradation and some high molecular weight contaminants in the 70-100 kDa range (possibly dimers or interactions with other proteins) can be observed in the samples with the highest ToTK1 content.

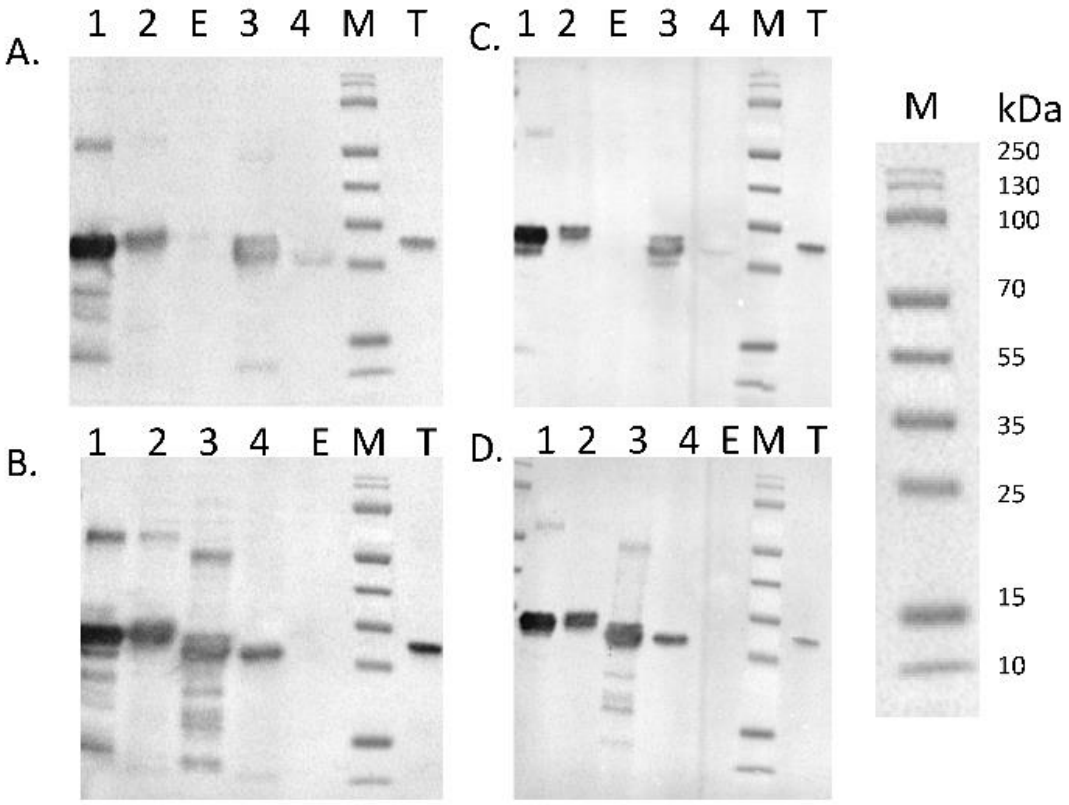

Figure 12 Western blot and immunodetection of ToTK1 expression in insect cells under defined MOI conditions. Similar amounts of protein (10 µg or 2 µg per lane) were applied for all cell extracts. To the right, the approximate sizes of the molecular weight markers (lane M) are indicated. Purified and His-tagged TEV protease (50 ng) was used as positive control for the blotting procedure and immunodetection, as well as internal standard. (A) High Five™ crude extract, 10 µg total protein. (B) Sf9, 10 µg. (C) High Five™, 2 µg. (D) Sf9, 2 µg. In all subfigures, the lanes are labeled as follows. Lane M, molecular weight marker. Lane T, TEV protease. Lane E, empty baculovirus (eGFP expression only). Lane 1, ToTK1-C-His. Lane 2, N-His-ToTK1. Lane 3, ToTK1ΔC25-C-His. Lane 4, N-His-ToTK1ΔC25

Discussion

Plant dNKs have been poorly studied and characterized so far (Clausen, 2012; Munch-Petersen, 2012). Here, we focused on a thymidine kinase from Tomato (*Solanum lycopersicum),* which play an important role in the salvage pathway for thymidine nucleotides synthesis during biochemical processes that are involved in various areas of plant physiology such as growth and development of plant embryos, germination and ripening of fruits (Satsolla, 2003).

Tomato Thymidine Kinase 1 (ToTK1) was first characterized in 2010. The gene coding for the wild-type ToTK1 (GeneBank ID: AF514755) was first cloned into a pGEX-2T vector and the recombinant plasmid was transformed into TK-deficient *E. coli* KY895 for expression (Khan 2010). In another recent study by Christiansen et al. in 2015, a mutant version of ToTK1 that has a 25-amino acid deletion at the C-terminus, was created based on the wt ToTK1 as a template, and also expressed in TK-deficient *E. coli* KY895. Expression of kinases in TK-deficient host gives several advantages: the lack of endogenous TKs eliminates the background kinase activities and as a result facilitates the measurement of ToTK1 activities. Moreover, it makes it possible to identify variants of ToTK1 with improved specificity towards NAs like AZT (Christiansen, 2015).

Choosing the expression vector

Unlike the previous studies for ToTK1 recombinant production, in this present study a mutant Tomato Thymidine Kinase 1 (**ToTK1ΔC25**) was cloned as His-tagged fusion proteins into a pET vector (instead of pGEX-2T) for expression in TK-deficient *E. coli* hosts with the aim of producing **high** amounts of pure protein suitable for crystallization. The choose of pET vector could be explained by the fact that expression of the target gene using a pET vector is controlled by the strong bacteriophage T7 promoter which produces high levels of transcription in *E. coli* using T7 RNA Polymerase (Studier & Moffatt, 1986). In comparison with the previous studies, expression of ToTK1 using pGEX vectors containing the weaker *tac* promoter resulted in production of lower amounts of recombinant protein (Khan 2010, Christiansen 2015).

Although expression using pET vectors produces high amounts of target protein, the produced recombinant protein had lower solubility and higher tendency to precipitate. A study carried out by Birringer *et al.* (2006) in attempt to express human thymidine kinase 1 (hTK1) using several pET vectors in various *E. coli* expression hosts, had resulted in production of only insoluble protein with minimal yields and poor quality (low activity) (Birringer, 2006).

The solubility of produced recombinant proteins is influenced by several parameters such as vector (promoter involved), expression host, cultivation conditions, temperature during expression and the concentration of inducer (IPTG). All those parameters contribute to the quality and quantity of the produced protein. However, protein misfolding and inclusion body formation are very common issues when a recombinant protein is overexpressed in *E. coli* (François Baneyx & Mirna Mujacic, 2004). One explanation for this pattern is that the production of recombinant protein using pET vector is occurred at a high rate that exceeds the folding capacities of the host. A very interesting study by Tegel *et al.* in 2011 had analyzed the amount and solubility of 17 different recombinant proteins produced in various *E. coli* strains under the control of different promoters (T7, *trc* and *lac*UV5). This study concluded that the use of the strong T7 promoter results in higher yields of total protein in most cases. However, the fraction of soluble protein was lower in case of using T7 promoter compared to the weaker *lac*UV5 promoter (Tegel, 2011).

Choosing the expression host

The choice of bacterial strain is another important factor that affects protein production. And that was clearly observed when looking to the results obtained from the expression trials. The difference in protein expression could be explained by the difference between the three *E. coli* strains used for expression. The *E. coli* Tuner™ strain its derivative Rosetta™ 2 are *lacY1* deletion mutants of BL21 that enable adjustable levels of protein expression throughout all cells in a culture. The absence of the *lac* permease allows a uniform entry of the inducer IPTG into all bacterial cells in the population and as a result, the level of expression can be regulated by adjusting the IPTG concentration in the growth medium (Novagen, 2003). Choosing Rosetta™ 2 strain as the best performing host for expression of the human-optimized ToTK1ΔC25 could be attributed to its interesting properties. The Rosetta™ 2 strain is designed to enhance the expression of eukaryotic proteins that contain codons **rarely** used in *E. coli.* Moreover, Rosetta™ 2 strain has additional plasmid (pRARE2) that supply tRNAs for the rare codons AUA, AGG, AGA, CUA, CCC and GGA. By supplying rare codons, the Rosetta strain provides for "universal" translation, where translation would otherwise be limited by the codon usage of E. coli (Novy et al., 2001).

The levels of expression in Rosetta™ strains have been investigated by Tegel *et al*. The study showed that Rosetta™ strains expressing a variety of proteins under the control of the T7 promoter were found to be more efficient at producing higher levels of total soluble protein than the standard BL21(DE3) strain due to their superior translation capabilities (Tegel, 2011). It is noteworthy that the ToTK1ΔC25 coding sequence used in the present study has been optimized for expression in *H. sapiens*, not in *E. coli*. The optimized sequence contains many codons that are rare in *E. coli*: 13 of them code for arginine (7 AGA and 6 CGG), 8 codes for proline (CCC), and 1 codes for glycine (GGA). This might explain why expression in Rosetta™ 2 is more favorable than in Tuner™ and ClearColi®.

The low levels of expression obtained might be explained by the C-terminal truncation which might affect the domain responsible for regulation of transcription and as a result affect the level of expression (Weston, 2014). Moreover, the location of the histidine tag whether it's fused to the N or C terminus might play a role in protein stability and activity. Since the 3D structure is not resolved yet, it's difficult to confirm the effect of the N-terminal histidine tag on the protein structural stability and activity. However, that might be an explanation for the absence of expression in the N-terminal his-tagged construct.

Maintaining good and controlled cultivation conditions is a crucial issue for better protein yield and quality. Induction temperature is a very important factor affecting protein yield and solubility. Although induction at high temperature increases the rate of protein synthesis but also leads to improper folding and generation of inclusion bodies. Therefore, it is not surprising that for all *E. coli* strains and constructs induced at a lower temperature at 25 °C resulted in higher yields of recombinant protein in the soluble fraction compared to the induction at 30°C. Temperatures as low as 14 °C over a period of 48 h have been used to produce large amounts of soluble human TK1 (Birringer, 2006). That explains why 25 °C has been chosen as the optimal temperature for expression. Previous studies had been reported choosing 25°C for the expression of other TK1-like kinases (Sandrini 2007, Carnrot, 2003; Sandrini, 2007b, Clausen, 2012; Khan 2010). On the other hand, the different concentrations of IPTG tested (0.4 and 0.8 mM, and later 0.2 mM) resulted in very similar protein expression profiles. A final concentration of 0.4–1.0 mM IPTG is recommended for full induction (Novagen, 2003).

Purification of ToTK1ΔC25 from *E. coli* by means of IMAC

As shown in figure (9), ToTK1ΔC25 was successfully purified with reasonable level of purity. That supports our choice of affinity chromatography which has several benefits. It is a single-step purification procedure that has high binding capacity and selectivity for the target protein. Moreover, the procedure is simple to conduct and robust (Schäfer 2015).

Purification of Tomato TK1 was previously performed using affinity chromatography based on GST tag. The methodology was for example described by Knecht *et al*. for the production and purification of *Drosophila melanogaster* deoxynucleoside kinase (*Dm*-dNK) mutants, which involved the overexpression of the constructs in *E. coli* as fusion proteins with an N-terminal GST tag. Purification was achieved by means of glutathione-Sepharose affinity chromatography followed by thrombin cleavage (Knecht 2007, Christiansen 2015).

Interestingly, in our present study ToTK1ΔC25 had been expressed as a His-tagged fusion protein for the first time (instead of GST) and purified by affinity chromatography using a column charged with Ni^{+2}. The use of histidine tag for purification is more preferable than GST tag. His tag is a short oligohistidine that has a length of 8 residues only compared to the 200 amino acid GST tag which has itself a size similar to the size of the produced protein (~26 kDa).

The production of a GST-tagged protein requires more resources than a His-tagged one and that could result in lower protein yields compared with the His-tagged protein. However, GST tags have been shown to improve the solubility of many eukaryotic proteins that are not soluble when expressed in *E. coli* (Smith and Johnson 1988). On the contrary, previous studies with human proteins produced in *E. coli* have shown that both N- and C-terminal His tags have a significant negative effect on protein solubility (Woestenenk 2004).

Among the strategies performed in order to improve the solubility of the His-tagged ToTK ΔC25, the tobacco etch virus (TEV) protease was used to cleave the oligohistidine tag from the purified protein. Although the cleavage reaction was performed successfully with reasonable efficiency, a considerable amount of the target protein got precipitated during the cleavage reaction that was performed overnight. That might be explained by the lower total protein yields obtained after TEV cleavage. However, the removal of the his tag resulted in a better protein solubility and stability compared to the uncleaved samples.

The additives used as an attempt to improve the solubility and stability of the purified protein indicate that ToTK1ΔC25 is more stable in the presence of dithiothreitol (DTT) and the zwitterionic detergent CHAPS as opposed to the nonionic surfactant Triton X-100 (Which promotes precipitation). Stability is also promoted by decreasing the concentration of NaCl (~130 mM) and imidazole (<30 mM) and by supplementation with substrates (ATP) and cofactors (Mg^{2+}). It has been reported that pre-incubation with 3.1 mM ATP produces a 1.7-fold increase in the Vmax value of ToTK1 (Mutahir, 2011). Thus, the presence of ATP might force the enzyme to adopt more active conformation that is less susceptible to denaturation.

The results obtained from the initial purification procedure had shown minor impurities and degradation of the target protein. Previous studies reported that truncated ToTK1 is more prone to degradation and contamination (Christiansen et al., 2015).

BEVS

Expression of recombinant proteins in insect cells using BEVS approach has become a popular eukaryotic expression system due to many advantages. In particular, the ability to produce high levels of recombinant protein with the appropriate post-translational modifications and improved quality (less intracellular degradation and better stability) (van Oers, 2011). Furthermore, BEVS is often the next expression system of choice if the protein expression in *E. coli* doesn't deliver the desired results.

Many studies had reported the use of BEVS as a powerful approach to produce high quality recombinant proteins required for structural studies. In 2012, 100 structures were determined and deposited in PDB from proteins produced mainly by the BEVS (Assenberg, 2013). The main goal of this project is to produce recombinant proteins, as pure and unmodified as possible, for 3D structure comparison of wild-type ToTK1 to a selected C-terminal truncated variant (ToTK1ΔC25). Therefore, BEVS has been chosen as a very good alternative expression system that might overcome the complications we had in *E. coli*.

Interestingly, a very recent publication by Timm *et al.* had reported the same complications that we experienced with the production of ToTK1 in *E. coli* for the expression of a TK1 from *Leishmania major* in *E .coli* (Timm, 2015). For successful production of material for structural determination of the L*. major* TK, these authors also switched to the BEVS. The methodology used by Timm *et al.* was very similar to our methodology with some variations. For the construction of baculoviruses, Timm *et al.* cloned different versions of *Lm*-TK (wild-type and C-terminal truncated) into transfer vectors that were later used to transfect Sf9 cells. Baculoviral stocks were amplified in a similar way as described in this study. However, protein expression was attempted in Sf9 cells **only**, and not in High Five™, which have been reported to consistently express more protein than Sf9 cells (Laukkanen, 1996). Noteworthy, the expression of Lm-TK was not performed under defined MOI conditions, but by using two viral concentrations (5 µL or 50 µL of virus per 5 mL of culture) and two harvest points (48 and 72 hours). However, in our present study, expression of ToTK1ΔC25 was performed under defined MOI in order to determine the optimal infection parameters to use for protein expression. In principle, a lower MOI (<1) is normally considered not infecting all cells in the culture directly and thus, takes longer time to give a product, while using a high MOI (from 5 and up) gives high probability to infect all the cells in the culture in a uniform manner and as a result gives more product (optimal expression) (Bac-to-Bac® Manual). Intracellular protein normally expresses best 48-72 h after uniform infection (using high MOI). In this study, we demonstrated the use of two MOI (0.5 and 5) for each construct/cell line, as a compromise to see which MOI gives the best expression in the respective cell line (WK personal communication).

Consequently, both wild-type and C-terminal truncated ToTK1 were successfully expressed in both Sf9 and High FiveTM insect cells under the defined conditions. Activity measurements for TK activity of cell extracts (**Table 1.)** clearly show an increase in detected TK activity compared to MOCK infected cells. For each tested construct and MOI, a higher specific activity was found in Sf9 cells compared to High Five™ cells. However, the total yield of activity in High Five™ exceeded yields in Sf9 by far. In addition, a MOI of 5 resulted in higher activity values, showing that a uniform infection works better than introducing a slight delay, as tested with a subhomogenous MOI of 0.5. Therefore expression of all constructs in High Five™ cells with a higher MOI, like 5, seems favorable to achieve the highest possible activity yield.

The mutant ToTK1ΔC25 has been reported to have a V_{max} value 1.4-fold higher than wild-type ToTK1 with deoxythymidine as substrate, and 1.2 fold increase in specificity towards AZT (Christiansen 2015). However, it is difficult to compare the expression levels of the different proteins with each other only based on the activity level, as they probably have different kinetic profiles. Therefore, we attempted visualization of the expression levels by western blotting.

Although it was possible to detect significant TK activities in the crude extracts of transfected Sf9 and Hi5 cells, expression scale-up and further purification of ToTK1ΔC25 has not been attempted yet. However, the findings of the present study pave the road to determine the 3D structure of ToTK1 and to better understanding of structure-function relationship.

Final Conclusions and future directions

In the present study, we have successfully cloned and expressed a C-terminal truncated version of ToTK1 in *E. coli* (Rosetta™ 2) under optimized conditions. The recombinant his-tagged protein (ToTK1ΔC25) was successfully purified from *E. coli* (Rosetta™ 2) by means of affinity chromatography (IMAC) which resulted in a pure sample with ~90% of homogeneity as judged by SDS-PAGE. Although it was possible to produce ToTK1ΔC25 in *E. coli* with relatively high yields of the recombinant protein, it wasn't possible to detect significant enzymatic TK activity in the purified samples or even in the crude extracts due to the high tendency of the expressed protein to form misfolded protein aggregates.

One suggestion to overcome the problem of protein precipitation and to achieve high yields of soluble and active enzyme, is to replace the oligohistidine tag with a GST fusion tag, which is known to improve solubility, and to maintain the pET expression system (making use of the T7-*lac* promoter). Protein folding might be improved by overexpressing chaperones in the prokaryotic host. We also suggest some modifications in the cultivation conditions such as, lower induction temperature and lower IPTG concentration.

We have also demonstrated that it is possible to successfully express ToTK1ΔC25 in both Sf9 and High Five™ insect cells by means of the baculovirus expression vector system (BEVS) with significant enzymatic activity and stability. Future directions should be aimed at scaling up ToTK1ΔC25 expression under the optimal conditions determined, as well as at performing purification trials of wt ToTK1 and and ToTK1ΔC25 from Sf9 and High Five™ insect cells followed by activity measurements and characterization of the purified proteins.

Finally, production of ToTK1ΔC25 using the BEVS has the potential to generate samples suitable for crystallization experiments, as demonstrated recently for the TK of *Leishmania major*. The results of this project are a major step on the road to determine the tertiary structure of the first TK1 of plant origin, and that will give better understanding of structure-function relationship and effective use of protein engineering methods to further improve the kinetic properties and stability of an enzyme with great potential for dNK-mediated suicide gene therapy for the treatment of malignant gliomas.

References

1. Aghi M, Hochberg F, Breakefield XO: **Prodrug activation enzymes in cancer gene therapy.** *J Gene Med* 2000, **2**:148–64.

2. Ardiani A, J. Johnson A, Ruan H, Sanchez-Bonilla M, Serve K, E. Black M: **Enzymes To Die For: Exploiting Nucleotide Metabolizing Enzymes for Cancer Gene Therapy**. *Curr Gene Ther* 2012, **12**:77–91.

3. Arnér ESJ, Eriksson S: **Mammalian deoxyribonucleoside kinases**. *Pharmacol Ther* 1995, **67**:155–186.

4. Assenberg R, Wan PT, Geisse S, Mayr LM: **Advances in recombinant protein expression for use in pharmaceutical research.** *Curr Opin Struct Biol* 2013, **23**:393–402.

5. Baneyx F, Mujacic M: **Recombinant protein folding and misfolding in *Escherichia coli***. *Nat Biotech* 2004, **22**:1399–1408.

6. Birringer MS, Claus MT, Folkers G, Kloer DP, Schulz GE, Scapozza L: **Structure of a type II thymidine kinase with bound dTTP.** *FEBS Lett* 2005, **579**:1376–82.

7. Birringer MS, Perozzo R, Kut E, Stillhart C, Surber W, Scapozza L, Folkers G: **High-level expression and purification of human thymidine kinase 1: quaternary structure, stability, and kinetics.** *Protein Expr Purif* 2006, **47**:506–15.

8. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding**. *Anal Biochem* 1976, **72**:248–254.

9. Carnrot C, Wehelie R, Eriksson S, Bölske G, Wang L: **Molecular characterization of thymidine kinase from Ureaplasma urealyticum: nucleoside analogues as potent inhibitors of mycoplasma growth**. *Mol Microbiol* 2003, **50**:771–780.

10. Carroni M, Kummer E, Oguchi Y, Wendler P, Clare DK, Sinning I, Kopp J, Mogk A, Bukau B, Saibil HR: **Head-to-tail interactions of the coiled-coil domains regulate ClpB activity and cooperation with Hsp70 in protein disaggregation**. *Elife* 2014, **3**:e02481.

11. Christiansen L, Egeblad L, Munch-Petersen B, Piškur J, Knecht W: **New Variants of Tomato Thymidine Kinase 1 Selected for Increased Sensitivity of *E. coli* KY895 towards Azidothymidine**. *Cancers (Basel)* 2015, **7**:966–980.

12. Ciccarone V, Polayes D, Luckow V: **Generation of Recombinant Baculovirus DNA in *E. coli* Using a Baculovirus Shuttle Vector**. In *Molecular Diagnosis of Infectious Diseases SE - 13*. *Volume 13*. Edited by Reischl U. Humana Press; 1998:213–235. [*Methods in Molecular Medicine™*]

13. Clausen AR, Girandon L, Ali A, Knecht W, Rozpedowska E, Sandrini MPB, Andreasson E, Munch-Petersen B, Piškur J: **Two thymidine kinases and one multisubstrate deoxyribonucleoside kinase salvage DNA precursors in Arabidopsis thaliana**. *FEBS J* 2012, **279**:3889–3897.

14. Clausen AR, Mutahir Z, Munch-Petersen B, Piškur J: **Plants Salvage Deoxyribonucleosides in Mitochondria**. *Nucleosides, Nucleotides and Nucleic Acids* 2014, **33**:291–295.

15. Fillat C, Carrió M, Cascante a, Sangro B: **Suicide gene therapy mediated by the Herpes Simplex virus thymidine kinase gene/Ganciclovir system: fifteen years of application.** *Curr Gene Ther* 2003, **3**:13–26.

16. Freeman SM, Abboud CN, Whartenby KA, Packman CH, Koeplin DS, Moolten FL, Abraham GN: **The "bystander effect": tumor regression when a fraction of the tumor mass is genetically modified.** *Cancer Res* 1993, **53**:5274–83.

17. Hartman SC, Buchanan JM: **Nucleic Acids, Purines, Pyrimidines (Nucleotide Synthesis)**. *Annu Rev Biochem* 1959, **28**:365–410.

18. Hébrard C, Cros-Perrial E, Clausen AR, Dumontet C, Piškur J, Jordheim LP: **Bacterial Deoxyribonucleoside Kinases are Poor Suicide Genes in Mammalian Cells**. *Nucleosides, Nucleotides and Nucleic Acids* 2009, **28**:1068–1075.

19. Huber BE, Austin EA, Richards CA, Davis ST, Good SS: **Metabolism of 5-fluorocytosine to 5-fluorouracil in human colorectal tumor cells transduced with the cytosine deaminase gene: significant antitumor effects when only a small percentage of tumor cells express cytosine deaminase**. *Proc Natl Acad Sci* 1994, **91** (17):8302–8306.

20. Invitrogen, (Life Technologies): **Bac-to-Bac® Baculovirus Expression System**. *Biotechnology* 2013.

21. Jordheim LP, Durantel D, Zoulim F, Dumontet C: **Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases**. *Nat Rev Drug Discov* 2013, **12**:447–464.

22. Khan Z, Knecht W, Willer M, Rozpedowska E, Kristoffersen P, Clausen AR, Munch-petersen B, Ekstro TJ: **Suicide Gene for Malignant Glioma Therapy**. *Neuro Oncol* 2010, **12**:549–558.

23. Kidd IM, Emery VC: **The use of baculoviruses as expression vectors.** *Appl Biochem Biotechnol* 1993, **42**:137–59.

24. King LA, Possee RD: ***The Baculovirus Expression System: A Laboratory Guide*.** Dordrecht: Springer Netherlands; 1992.

25. Knecht W, Petersen GE, Munch-Petersen B, Piskur J: **Deoxyribonucleoside kinases belonging to the thymidine kinase 2 (TK2)-like group vary significantly in substrate specificity, kinetics and feed-back regulation.** *J Mol Biol* 2002, **315**:529–40.

26. Knecht W, Petersen GE, Sandrini MPB, Søndergaard L, Munch-Petersen B, Piškur J: **Mosquito has a single multisubstrate deoxyribonucleoside kinase characterized by unique substrate specificity**. *Nucleic Acids Res* 2003, **31**:1665–1672.

27. Konrad Rajab PNEZK and CC: **Suicide Gene Therapy against Cancer**. *J Genet Syndr Gene Ther* 2013, **2013**.

28. Konrad A, Lai J, Mutahir Z, Piškur J, Liberles D: **The Phylogenetic Distribution and Evolution of Enzymes Within the Thymidine Kinase 2-like Gene Family in Metazoa**. *J Mol Evol* 2014, **78**:202–216.

29. Laukkanen ML, Oker-Blom C, Keinänen K: **Secretion of green fluorescent protein from recombinant baculovirus-infected insect cells.** *Biochem Biophys Res Commun* 1996, **226**:755–61.

30. Lawler SE, Peruzzi PP, Chiocca E a: **Genetic strategies for brain tumor therapy.** *Cancer Gene Ther* 2006, **13**:225–233.

31. Louis DN: **Molecular pathology of malignant gliomas.** *Annu Rev Pathol* 2006, **1**:97–117.

32. Määttä A-M, Samaranayake H, Pikkarainen J, Wirth T, Ylä-Herttuala S: **Adenovirus mediated herpes simplex virus-thymidine kinase/ganciclovir gene therapy for resectable malignant glioma.** *Curr Gene Ther* 2009, **9**:356–67.

33. Moolten FL: **Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy.** *Cancer Res* 1986, **46**:5276–81.

34. Moolten FL, Wells JM: **Curability of Tumors Bearing Herpes Thymidine Kinase Genes Transfered by Retroviral Vectors**. *J Natl Cancer Inst* 1990, **82** (4):297–300.

35. Munch-Petersen B, Knecht W, Lenz C, Søndergaard L, Piškur J: **Functional Expression of a Multisubstrate Deoxyribonucleoside Kinase from Drosophila melanogaster and Its C-terminal Deletion Mutants** . *J Biol Chem* 2000, **275** (9):6673–6679.

36. Munch-Petersen B, Piskur J, Søndergaard L: **Four Deoxynucleoside Kinase Activities from Drosophila melanogaster Are Contained within a Single Monomeric Enzyme, a New Multifunctional Deoxynucleoside Kinase** . *J Biol Chem* 1998, **273** (7):3926–3931.

37. Munch-Petersen B, Piskur J, Søndergaard L: **The Single Deoxynucleoside Kinase in Drosophila Melanogaster, Dm-dnk, is Multifunctional and Differs from the Mammalian Deoxynucleoside Kinases**. In *Purine and Pyrimidine Metabolism in Man IX SE - 92*. *Volume 431*. Edited by Griesmacher A, Müller M, Chiba P. Springer US; 1998:465–469. [*Advances in Experimental Medicine and Biology*]

38. Mutahir Z, Clausen AR, Andersson KM, Wisen SM, Munch-Petersen B, Piškur J: **Thymidine kinase 1 regulatory fine-tuning through tetramer formation**. *FEBS J* 2013, **280**:1531–1541.

39. Novagen: **pET System Manual Novagen pET System Manual Novagen**. *Biosystems* 2003:1–68.

40. Novy R, Drott D, Yaeger K, Mierendorf R: **Overcoming the codon bais of *E. coli* for enhanced protein expression**. *Innovations* 2001:4–6.

41. Omuro A: **Glioblastoma and Other Malignant Gliomas**. *Jama* 2013, **310**:1842.

42. Osaki T, Tanio Y, Tachibana I, Hosoe S, Kumagai T, Kawase I, Oikawa S, Kishimoto T: **Gene therapy for carcinoembryonic antigen-producing human lung cancer cells by cell type-specific expression of herpes simplex virus thymidine kinase gene.** *Cancer Res* 1994, **54**:5258–61.

43. Reichard P: **Interactions Between Deoxyribonucleotide and DNA Synthesis**. *Annu Rev Biochem* 1988, **57**:349–374.

44. Sandrini MPB, Söderbom F, Mikkelsen NE, Piškur J: **Dictyostelium discoideum Salvages Purine Deoxyribonucleosides by Highly Specific Bacterial-like Deoxyribonucleoside Kinases**. *J Mol Biol* 2007, **369**:653–664.

45. Sandrini MPB, Clausen AR, On SLW, Aarestrup FM, Munch-Petersen B, Piškur J: **Nucleoside analogues are activated by bacterial deoxyribonucleoside kinases in a species-specific manner**. *J Antimicrob Chemother* 2007, **60**:510–520.

46. Sandrini MPB, Piškur J: **Deoxyribonucleoside kinases: two enzyme families catalyze the same reaction**. *Trends Biochem Sci* 2015, **30**:225–228.

47. Schäfer F, Seip N, Maertens B, Block H, Kubicek J: **Purification of GST-Tagged Proteins.** *Methods Enzymol* 2015, **559**:127–39.

48. Segura-Peña D, Lutz S, Monnerjahn C, Konrad M, Lavie A: **Binding of ATP to TK1-like enzymes is associated with a conformational change in the quaternary structure.** *J Mol Biol* 2007, **369**:129–41.

49. Skovgaard T, Uhlin U, Munch-Petersen B: **Comparative active-site mutation study of human and Caenorhabditis elegans thymidine kinase 1**. *FEBS J* 2012, **279**:1777–1787.

50. Slot Christiansen L, Munch-Petersen B, Knecht W: **Non-Viral Deoxyribonucleoside Kinases-Diversity and Practical Use.** *J Genet Genomics* 2015, **42**:235–48.

51. Smith DB, Johnson KS: **Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase**. *Gene* 1988, **67**:31–40.

52. Stasolla C, Katahira R, Thorpe TA, Ashihara H: **Purine and pyrimidine nucleotide metabolism in higher plants.** *J Plant Physiol* 2003, **160**:1271–95.

53. Stedt H, Samaranayake H, Kurkipuro J, Wirth G, Christiansen LS, Vuorio T, Määttä a-M, Piškur J, Ylä-Herttuala S: **Tomato thymidine kinase-based suicide gene therapy for malignant glioma—an alternative for Herpes Simplex virus-1 thymidine kinase**. *Cancer Gene Ther* 2015, **22**:130–137.

54. Summers MD: **Milestones leading to the genetic engineering of baculoviruses as expression vector systems and viral pesticides.** *Adv Virus Res* 2006, **68**:3–73.

55. Tegel H, Ottosson J, Hober S: **Enhancing the protein production levels in *Escherichia coli* with a strong promoter**. *FEBS J* 2011, **278**:729–739.

56. Thelander L, Reichard P: **Reduction of Ribonucleotides**. *Annu Rev Biochem* 1979, **48**:133–158.

57. Timm J, Bosch-Navarrete C, Recio E, Nettleship JE, Rada H, González-Pacanowska D, Wilson KS: **Structural and Kinetic Characterization of Thymidine Kinase from Leishmania major**. *PLoS Negl Trop Dis* 2015, **9**:e0003781.

58. Tinta T, Christiansen LS, Konrad A, Liberles D., Turk V, Munch-Petersen B, Piškur J, Clausen AR: **Deoxyribonucleoside kinases in two aquatic bacteria with high specificity for thymidine and deoxyadenosine**. *FEMS Microbiol Lett* 2012, **331**:120–127.

59. Van Oers MM: **Opportunities and challenges for the baculovirus expression system**. *J Invertebr Pathol* 2011, **107**(SUPPL.):S3–S15.

60. Welin M, Kosinska U, Mikkelsen NE, Carnrot C, Zhu C, Wang L, Eriksson S, Munch-Petersen B, Eklund H: **Structures of thymidine kinase 1 of human and mycoplasmic origin.** *Proc Natl Acad Sci U S A* 2004, **101**:17970–17975.

61. Whitaker WR, Lee H, Arkin AP, Dueber JE: **Avoidance of Truncated Proteins from Unintended Ribosome Binding Sites within Heterologous Protein Coding Sequences.** *ACS Synth Biol* 2014.

62. Woestenenk E, Hammarström M, van den Berg S, Härd T, Berglund H: **His tag effect on solubility of human proteins produced in *Escherichia coli:* a comparison between four expression vectors**. *J Struct Funct Genomics* 2004, **5**:217–229.

63. Xu J, Zhang L, Yang D-L, Li Q, He Z: **Thymidine kinases share a conserved function for nucleotide salvage and play an essential role in Arabidopsis thaliana growth and development**. *New Phytol* 2015:n/a–n/a.

64. Xu Y-Z, Huang P, Plunkett W: **Functional Compartmentation of dCTP Pools: PREFERENTIAL UTILIZATION OF SALVAGED DEOXYCYTIDINE FOR DNA REPAIR IN HUMAN LYMPHOBLASTS** . *J Biol Chem* 1995, **270** (2):631–637.

65. Young JD, Yao SYM, Baldwin JM, Cass CE, Baldwin SA: **The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29.** *Mol Aspects Med* 2013, **34**:529–47.

66. Zhang Y, Morar M, Ealick SE: **Structural biology of the purine biosynthetic pathway**. *Cell Mol Life Sci* 2008, **65**:3699–3724.

67. Studier FW, Moffatt BA: **Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes**. *J Mol Biol* 1986, **189**:113–130.

Acknowledgments

It has been a long and eventful journey that comes to an end. A test of patience and devotion that I could not have passed without the help and support of my colleagues, friends and family. I would like to acknowledge the financial support provided by the Swedish Institute (SI) through the generous SI study scholarship during my whole study period at Lund University. I also would like to acknowledge the unlimited support and guidance provided by my supervisors, Louise Slot Christiansen and Wolfgang Knecht, as well as, my examiner Leif Bulow for the guidance and advice provided during my planning for this project. I also would like to thank Lund University staff members and student counselors for their generous help and guidance during my studies.

Finally, I would like to thank all of my friends and colleagues who supported and encouraged me. Special thanks to my colleague and friend Juan Manuel Orozco Rodriguez for his help and support during this project, it's been really a pleasure to work with you and discuss together many topics.

Appendix I. Materials

Chemical Reagents

Reagent	Manufacturer	Catalog number
Agarose	Sigma-Aldrich	Cat. No. A9539
Bacto-agar	Saveen Werner	Cat. No. B1000
Peptone (Trypton)	Merck Millipore	Cat. No. 107213
Yeast Extract	Fluka	Cat. No. 70161
Sodium Chloride	Merck Millipore	Cat. No. 106406
Sodium Fluoride	Sigma-Aldrich	Cat. No. 201154
Sodium hydroxide	Merck Millipore	Cat. No. 106462
Potassium Chloride	VWR	Cat. No. BDH9258
Magnesium chloride hexahydrate	Merck Millipore	Cat. No. 442615
Ammonium Formate	Fluka	Cat. No. 17843
Imidazole	Merck Millipore	Cat. No. 104716
Bradford Reagent	Bio-Rad	Cat. No. B6916
Biosafe Coomassie Stain	Bio-Rad	Cat. No. 1610786
Adenosine 5′-triphosphate disodium salt hydrate	Sigma-Aldrich	Cat. No. A26209
Potassium phosphate dibasic trihydrate	Sigma-Aldrich	Cat. No. P9666
Potassium phosphate monobasic, anhydrous	Sigma-Aldrich	Cat. No. P0662
Ethydium Bromide	Amresco	Cat. No. 0492
Isopropyl thiogalactoside (IPTG)	VWR	Cat. No. IC102101
cOmplete™ EDTA-free Protease Inhibitor Cocktail Tablets	Roche	Cat. No. 05892791001
Tris/Glycine Buffer (10×)	Bio-Rad	Cat. No. 1610734
Tris base	Merck Millipore	Cat. No. 648310
Hydrochloric Acid	Merck Millipore	Cat. No. 100317
[Methyl-^{3}H]-Thymidine (20Ci/mmol)	Perkin Elmer	Cat. No. NET027E005MC
Regular Thymidine (unlabeled)	Sigma-Aldrich	Cat. No. T9250
Dithiothreitol (DTT)	Sigma-Aldrich	Cat. No. D9779
Triton X-100	Sigma-Aldrich	Cat. No. X100

CHAPS (cholamidopropyl dimethylammonio propanesulfonate)	Sigma-Aldrich	Cat. No. C3023
EDTA	Sigma-Aldrich	Cat. No. E9884
Virkon™ Tablets	DuPont	-
Bluo-Gal	Sigma-Aldrich	Cat. No. B2904
DMSO (dimethyl sulfoxide)	Sigma-Aldrich	Cat. No. D8418
Tetracycline hydrochloride	Sigma-Aldrich	Cat. No. T3383
Chloramphenicol	Sigma-Aldrich	Cat. No. C0378
Gentamycin solution	Sigma-Aldrich	Cat. No. G1397
Kanamycin A sulfate	Sigma-Aldrich	Cat. No. K1637
Glycerol	Sigma-Aldrich	Cat. No. G5516
Ethanol	VWR	Cat. No. 20824
2-Propanol (Isopropanol)	Sigma-Aldrich	Cat. No. 278475
Mini-PROTEAN® TGX™ Precast Gels (12% polyacrylamide)	Bio-Rad	Cat. No. 4561043
EX-CELL® 420 Serum-Free Medium	Sigma-Aldrich	Cat. No. 14420C
HyClone SFX-Insect™ Serum-Free Medium	Thermo Fisher Scientific	Cat. No. SH30278.02
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	Cat. No. K0502
GeneJET PCR Purification Kit	Thermo Fisher Scientific	Cat. No. K0701
GeneJET Gel Extraction Kit	Thermo Fisher Scientific	Cat. No. K0691
PureLink® HiPure Plasmid Miniprep Kit	Thermo Fisher Scientific	Cat. No. K2100-02
Ultima Gold™ liquid scintillation fluid	Perkin Elmer	Cat. No. 6013329
FlashFECTIN™	Oxford Expression Technologies	Cat. No. 300104
SuperSignal™ West Pico Chemiluminescent Substrate	Thermo Fisher Scientific	Cat. No. 34080

Biological Materials

Strain/Cell line	Manufacturer	information
One Shot® TOP10 *E. coli* Competent Cells	Thermo Fisher Scientific	Amplification of pET-24a(+) vector and expression plasmids
ClearColi® BL21(DE3) *E. coli*	Lucigen	ToTK1ΔC25 expression host
Rosetta™ 2 (DE3) *E. coli*	Novagen	ToTK1ΔC25 expression host
Tuner™(DE3) *E. coli*	Novagen	ToTK1ΔC25 expression host
MAX Efficiency® DH10Bac™ *E. coli* Competent Cells	Thermo Fisher Scientific	Production of recombinant bacmids
Sf9 cells (*Spodoptera frugiperda*)	Thermo Fisher Scientific	Baculovirus amplification and ToTK1ΔC25 expression
High Five™ (Hi5) cells (*Trichoplusia ni*)	Thermo Fisher Scientific	ToTK1ΔC25 expression
Plasmids	**Manufacturer**	**Information**
pET-24a(+)	Novagen	5310 bp, pBR322 ori, T7-lac promoter, Kan^R
pDual-HoHis2	Provided by the supervisor	Expression plasmid carrying ToTK1ΔC25 with a C-terminal His tag
pGEX- ToTK1ΔC25	Provided by the supervisor	pGEX-2T expression plasmid bearing ToTK1ΔC25
pFastBac™ Dual GFP	Provided by the supervisor	Donor vector for expression of proteins in insect cells bearing GFP
FlashBAC™ DNA	Oxford Expression Technologies	Cat. No. 100151
T4 DNA ligase (5 Weiss U/μL)	Thermo Fisher Scientific	Cat. No. EL0014
DNA Taq Polymerase	Thermo Fisher Scientific	
Phusion High Fidelity DNA Polymerase (2000 U/ml)	New England Biolabs	Cat. No. M0530S
FastDigest™ restriction enzymes	Thermo Fisher Scientific	*NdeI* (Cat. No. FD0583) *NotI* (Cat. No. FD0595)
PageRuler Plus Prestained Protein Ladder (10−250 kDa)	Thermo Fisher Scientific	Cat. No. 26619
GeneRuler 1 kb DNA ladder	Thermo Fisher Scientific	Cat. No. SM0313
Mouse Anti-6×His	BD Biosciences	Cat. No. 552565
Goat Anti-Mouse IgG - Peroxidase	Sigma-Aldrich	Cat. No. A4416

Equipments and other consumables

Equipment/Consumable	Manufacturer	Model
Autoclave	-	-
Centrifuge (large capacity)	Beckman Coulter	Avanti J-20 XP (JA 25.50 and JLA 8.1000 rotors)
Microcentrifuge	Heraeus	Biofuge Pico
	Eppendorf	5424 R
Chromatography System	GE Healthcare	ÄKTA Explorer
Peristaltic pump	GE Healthcare	P-1
HisTrap HP (Ni^{2+} Sepharose column)	GE Healthcare	1 mL, 7×25 mm
Disposable Desalting Columns	GE Healthcare	PD-10
Electroporation device	Bio-Rad	Gene Pulser II with Pulse Controller Plus
Electroporation Cuvettes	Eppendorf	
Electrophoresis power supply	VWR	Power source 300 V
Electrophoresis cell	Bio-Rad	Mini-PROTEAN® Tetra Cell
Blotting System	Bio-Rad	Trans-Blot® Turbo™ Transfer System
Protein imaging system	Bio-Rad	ChemiDoc MP
Bioimaging system (DNA)	UVP	BioDoc-It 210
Fluorescence inverted microscope	Nikon	Diaphot (coupled to an Olympus DP70 camera)
Optical microscope	LRI Instrument AB	-
French Press	SLM Aminco	French® FA-032
Liquid Scintillation Counter	Perkin Elmer	Tri-Carb 2800 TR
Plastic scintillation vials	Perkin Elmer	20 mL
Spectrophotometer	Jenway	6705 UV/Vis
Spectrophotometer	Thermo Scientific	Nanodrop 2000c
Thermocycler (PCR)		
Sonicator	Sonics and Materials, Inc	Vibra Cell VC50
Centrifugal evaporator	Savant	SpeedVac SVC 100H
Incubation shaker (insect cells)	Infors HT	Multitron Pro and Multitron Standard
Incubation shaker (bacterial cells)	New Brunswick Scientific	Innova 4300
Milli-Q Apparatus	Millipore Corporation	
Orbital shaker	Brinkmann	Heidolph Unimax 2010
Water bath	Heto	OBN 18 HMT 100

pH-meter	Mettler Toledo	Seven Multi
Serological pipettes	Sarstedt	
Microcentrifuge tubes	Eppendorf	
TPP TubeSpin® Bioreactors	TPP Techno Plastic Products AG	
Vivaspin® 20 Centrifugal concentrator (20 mL, 3 kDa)	GE Healthcare	Cat. No. VS2091
PD-10 Desalting column	GE Healthcare	
Conical tubes	Sarstedt	
Syringe Filters	Sarstedt	
Microfilters (0.45 µm, cellulose acetate)	Whatman	
DE-81 DEAE ion exchange paper	Whatman	
Erlenmeyer Flasks	Sigma-Aldrich	Corning

Primers

Generation of N-terminal His-tagged ToTK1ΔC25 constructs

Forward Primer 1 (P1F)

5′-GCGGATCCCATATGCACCATCATCACCACCATCACCATTCCAGCGGTGTGG-3′

Forward Primer 2 (P2F)

5′-CATCACCATTCCAGCGGTGTGGATCTCGGCACCGAGAACTTGTACTTCC-3′

Forward Primer 3 (P3F)

5′-CGGCACCGAGAACTTGTACTTCCAGTCAATGGCCTTCAGCAGCAGCGCC-3′

Reverse Primer 2 (P2R)

5′-CGCGGCCGCTTATTACACGCTCTGGCCGTTCACGTAG-3′

PCR analysis of recombinant bacmid

pUC/M13 Forward Primer

5′-CCCAGTCACGACGTTGTAAAACG-3′

pUC/M13 Reverse Primer

5′-AGCGGATAACAATTTCACACAGG-3′

Appendix II. Media and Buffers

LB Media (per 1 L)

NaCl	4 g
Peptone	10 g
Yeast Extract	5 g

Adjust pH to 7.0 with NaOH
(add 15 g Bacto-agar for solid media)

TB Media (per 1 L)

Nutrient solution (900 mL)

Peptone	12 g
Yeast Extract	24 g
Glycerol	4 mL

K-phosphate solution (100 mL)

K_2HPO_4	12.54 g
KH_2PO_4	2.31 g

Autoclave separately and combine both solutions

Lysis Buffer (Bacterial Cells)

Tris-HCl pH 7.5	50 mM
NaCl	300 mM
Imidazole	20 mM
DTT	1 mM
Protease Inhibitor Mix	1X (added right before use)

Lysis Buffer (Insect Cells)

Tris-HCl pH 7.5	50 mM
NaCl	300 mM
CHAPS	1 mM
DTT	1 mM
Protease Inhibitor Mix	1X (added right before use)

Antibiotics (working concentrations)

Kanamycin	50 µg/mL
Chloramphenicol	34 µg/mL
Tetracycline	10 µg/mL
Gentamicin	7 µg/mL

Chromatography Buffers

Binding Buffer (BB)

Tris-HCl pH 8.0	50 mM
NaCl	300 mM
Imidazole	20 mM

Elution Buffer (EB)

Tris-HCl pH 8.0	50 mM
NaCl	300 mM
Imidazole	300 mM

Dilution Buffer (DB)

Tris-HCl pH 8.0	50 mM
NaCl	100 mM
DTT	1 mM

Imidazole Wash Buffer (WB)

Tris-HCl pH 8.0	50 mM
NaCl	300 mM
Imidazole	500 mM

ATP Wash Buffer

Binding Buffer + 10 mM ATP + 10 mM $MgCl_2$

Buffers for Enzyme Assays

Enzyme dilution buffer

Tris-HCl, pH 7.5	50 mM
CHAPS	1 mM
DTT	5 mM
BSA	3 mg/mL

ATP/MgCl2 Mix (10×)

Tris-HCl, pH 7.5	500 mM
DTT	100 mM
ATP	25 mM
$MgCl_2$	25 mM

Enzyme assay mix

^{3}H-methyl thymidine*	100 µM
BSA	3 mg/mL
CHAPS	0.5 mM
ATP/$MgCl_2$ Mix	1×

*Diluted to 200 µCi/µmol with unlabeled thymidine

dTMP Elution Solution

KCl	0.2 M
HCl	0.1 M

Buffers for Western Blotting

PBS (10×)

NaCl	1.4 M
KCl	27 mM
Na_2HPO_4	101 mM
KH_2PO_4	18 mM

pH is adjusted to 7.3

Transfer Buffer (per 1 L)

Bio-Rad 10× Tris/Glycine Buffer	100 mL
Ethanol	200 mL
ddH_2O	700 mL

Appendix III. DNA and protein sequences

Coding DNA Sequence (CDS) of the C-terminal truncated ToTK1 (**ToTK1ΔC25**). The sequence is codon-optimized for expression in *H. Sapiens*

```
  1 ATGGCCTTCA GCAGCAGCGC CAGAAACCCC GTGGACCTGC
 41 GGAACGGCAG CAAGAACAGC TTCTGCCCCG TGGGCGAGAT
 81 CCACGTGATC GTGGGCCCCA TGTTCGCCGG CAAGACCACC
121 GCCCTGCTGC GGAGAGTGAA CCTGGAAAGC AACGACGGCC
161 GGAACGTGGT GCTGATCAAG AGCAGCAAGG ACGCCAGATA
201 CGCCGTGGAT GCTGTGGTGA CACACGACGG CACCCGGTTC
241 CCTTGCTGGT CCCTGCCCGA CCTGAGCAGC TTCAAGCAGA
281 GATTCGGCAA GGATGCCTAC GAGAAGGTGG ACGTGATCGG
321 CATCGACGAG GCCCAGTTCT TCGGCGACCT GTACGAGTTC
361 TGCTGCAACG CCGCCGACTT CGACGGCAAG ATCATCGTGG
401 TGGCCGGCCT GGACGGCGAC TACCTGAGAA AGTCCTTCGG
441 CAGCGTGCTG GACATCATCC CCCTGGCCGA CACCGTGACC
481 AAGCTGACCG CCAGATGCGA GCTGTGCAAC AGACGGGCCT
521 TCTTCACCTT CCGCAAGACC AACGAGACAG AGACAGAGCT
561 GATCGGCGGA GCCGACATCT ACATGCCCGT GTGCCGGCAG
601 CACTACGTGA ACGGCCAGAG CGTGTGA
```

Protein Sequence of the C-terminal truncated ToTK1 (**ToTK1ΔC25**). This protein sequence was obtained by back translation of the CDS of ToTK1ΔC25 using ***EMBOSS Transeq Tool***

```
  1 MAFSSSARNP VDLRNGSKNS FCPVGEIHVI VGPMFAGKTT
 41 ALLRRVNLES NDGRNVVLIK SSKDARYAVD AVVTHDGTRF
 81 PCWSLPDLSS FKQRFGKDAY EKVDVIGIDE AQFFGDLYEF
121 CCNAADFDGK IIVVAGLDGD YLRKSFGSVL DIIPLADTVT
161 KLTARCELCN RRAFFTFRKT NETETELIGG ADIYMPVCRQ
201 HYVNGQSV*
```

The Coding DNA Sequence (CDS) of the **wild-type ToTK1**, optimized for expression in *H. sapiens* (GenBank Accession Number is AF514775.1)

```
  1 ATGGCCTTCA GCAGCAGCGC CAGAAACCCC GTGGACCTGC
 41 GGAACGGCAG CAAGAACAGC TTCTGCCCCG TGGGCGAGAT
 81 CCACGTGATC GTGGGCCCCA TGTTCGCCGG CAAGACCACC
121 GCCCTGCTGC GGAGAGTGAA CCTGGAAAGC AACGACGGCC
161 GGAACGTGGT GCTGATCAAG AGCAGCAAGG ACGCCAGATA
201 CGCCGTGGAT GCTGTGGTGA CACACGACGG CACCCGGTTC
241 CCTTGCTGGT CCCTGCCCGA CCTGAGCAGC TTCAAGCAGA
281 GATTCGGCAA GGATGCCTAC GAGAAGGTGG ACGTGATCGG
321 CATCGACGAG GCCCAGTTCT TCGGCGACCT GTACGAGTTC
361 TGCTGCAACG CCGCCGACTT CGACGGCAAG ATCATCGTGG
401 TGGCCGGCCT GGACGGCGAC TACCTGAGAA AGTCCTTCGG
441 CAGCGTGCTG GACATCATCC CCCTGGCCGA CACCGTGACC
481 AAGCTGACCG CCAGATGCGA GCTGTGCAAC AGACGGGCCT
521 TCTTCACCTT CCGCAAGACC AACGAGACAG AGACAGAGCT
561 GATCGGCGGA GCCGACATCT ACATGCCCGT GTGCCGGCAG
601 CACTACGTGA ACGGCCAGAG CGTGAACGAG AGCGCCAAGA
641 TGGTGCTGGA AAGCCACAAG GTGTCCAACG AGCTGATCCT
681 GGAATCTCCC CTGGTGGACC CCTGA
```

Amino acid sequence of the **wild-type ToTK1** (Uniprot ID: Q71F77, GenBank ID: AAQ08180.1)

```
  1 MAFSSSARNP VDLRNGSKNS FCPVGEIHVI VGPMFAGKTT
 41 ALLRRVNLES NDGRNVVLIK SSKDARYAVD AVVTHDGTRF
 81 PCWSLPDLSS FKQRFGKDAY EKVDVIGIDE AQFFGDLYEF
121 CCNAADFDGK IIVVAGLDGD YLRKSFGSVL DIIPLADTVT
161 KLTARCELCN RRAFFTFRKT NETETELIGG ADIYMPVCRQ
201 HYVNGQSVNE SAKMVLESHK VSNELILESP LVDP
```