NTNU – Trondheim
Norwegian University of
Science and Technology

Purification and characterization of an unusual DNA glycosylase in diatoms

Martin Vejle Andersen

Cell biology

Supervisor: Atle M. Bones, IBI
Co-supervisor: Tore Brembu, IBI
Geir Slupphaug, Institutt for kreftforskning og molekylær medisin

Norwegian University of Science and Technology
Department of Biology

Acknowledgements

The work of the master was performed at the Department of Biology and at the Department of Cancer Research and Molecular Medicine at The Norwegian University of Science and Technology (NTNU) from 2010 to 2012.

I would like to thank my supervisor professor Atle M. Bones, my academic supervisor researcher Tore Brembu and my co-supervisor professor Geir Slupphaug for believing in me, for their support and for sharing their great knowledge. I would also like to give a special thanks to Dr. Berit Doseth and principal engineer Sissel Håvåg for their work and assistance.

I am grateful to the members of the CMB group and other master students, especially Post doc. Ralph Kissen, Dr. Christopher Sørmo and Ph.D student Bjørnar Sporsheim for advice and for taking the time to help me.

Summary

DNA damage may be caused by many different agents, including radiation, spontaneous mutations and chemical mutagens. If the damage is left unrepaired, it may result in mutations or cell death. All organisms contain repair system to repair these damages. The mechanism that is believed to be the most common is base excision repair (BER). The BER pathway is initiated by enzymes called DNA glycosylases, which recognize and remove modified or misincorporated bases.

Studies of the sequenced genome of the two diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* have shown that they have an unusual DNA glycosylase. The glycosylase has been named “Dual DNA glycosylase” and has only been found in diatom genomes. DDG contain an N-terminal NEIL (endonuclease VIII/Nei like) domain and a C-terminal UNG (Uracil N-glycosylase) domain, connected with a linker region. DDG has later been found in the pennate diatoms *Fragilariopsis cylindrus* and *Seminavis robusta,* suggesting that DDG is a diatom-specific protein.

cDNA clones encoding the FL (full-length) *P. tricornutum* DDG as well as the single NEIL domain and UNG domain were cloned into the expression vector P-BADM-30, which contains both a histidine-tag and a glutathion S-transferase (GST)-tag for affinity purification. The clones were verified by sequencing and transformed into *Escerichia coli* ArcticExpress RIL cells. The ArcticExpress cells are modified for fusion protein expression. The expression of fusion proteins was optimized by induction time, temperature and inducer concentration. Expressed proteins were purified by the use of both histidin tag and GST tag columns, and verified by MALDI-TOF analysis.

Purified recombinant PtDDG-FL was used in UDG activity assay analysis to determine activity optimum. The pH optimum was found to be at pH 8, NaCl optimum at 125 mM, $MgCl_2$ optimum at 8 mM $MgCl_2$ and the temperature optimum at 45°C. PtDDG-FL substrate preference was found to be in the order of ssU>U:G>U:A. The thermal stability of PtDDG-FL at its optimum temperature was found to be low in comparison to human UNG2. The K_m and K_{cat} values of PtDDG-FL against double-stranded uracil-containing DNA were 0.64 µM and 3.3 min^{-1}, respectively. The corresponding values for the nuclear human UNG2 enzyme are 3.0 µM and 187 min^{-1}, respectively.

Table of Contents

Abberivations

$[^3H]$	Tritium
A	Adenine
AA	Amino acid
AID	Activation-induced cytosine deaminase
AP	Apurinic/apyrimidinic
APS	Ammonium persulphate
BER	Base excision repair
BSA	Bovine serum albumin
C	Cytosine
C-terminal	Carboxyl terminal
DDG	Dual DNA glycosylase
DNA	Deoxynucleotide triphosphate
ds	Double stranded
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
FL	Full-length
G	Guanine
H2TH	Helix-two-turn-helix motif
HSV	herpes simplex virus
Ig	Immunoglobulin
kDa	Kilodalton
LB	Luria-Bertani
LP	Long-patch
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MilliQ	water Ultra pure water
MWCO	Molecular weight cut off
Nei	Endonuclease VIII
NEIL	Endonuclease VIII-like
N-terminal	Amino terminal
Nth	Endonuclease III
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
pI	Isoelectric point
psi	Pounds per square inch
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SP	Short-patch
ss	Single-stranded
T	Thymine

TAE	Tris Acetate Ethylenediaminetetraacetic acid
TCA	Trichloroacetic acid
Tg	Thymine glycol
U:A	Uracil base paired with adenine
U:G	Uracil base paired with guanine
U	Uracil
UNG	Uracil N-glycosylase
UV	Ultraviolet
v/v	Volume/volume
w/v	Weight/volume

Theory

1.1 Diatoms

Diatoms are unicellular algae and a common form of phytoplankton. They can be found in the ocean, freshwater and in soils. Diatoms are thought to be the most important group of eukaryotic phytoplankton and are estimated to contribute to as much as one quarter of the global primary productivity in the ocean. There are more than 250 known genera of diatoms, with an estimate of about 100 000 species. A well-known characteristic of diatoms is their ability to generate a highly ornamented silica cell wall; known as the frustules (Figure 1.1) (Falciatore and Bowler 2002).

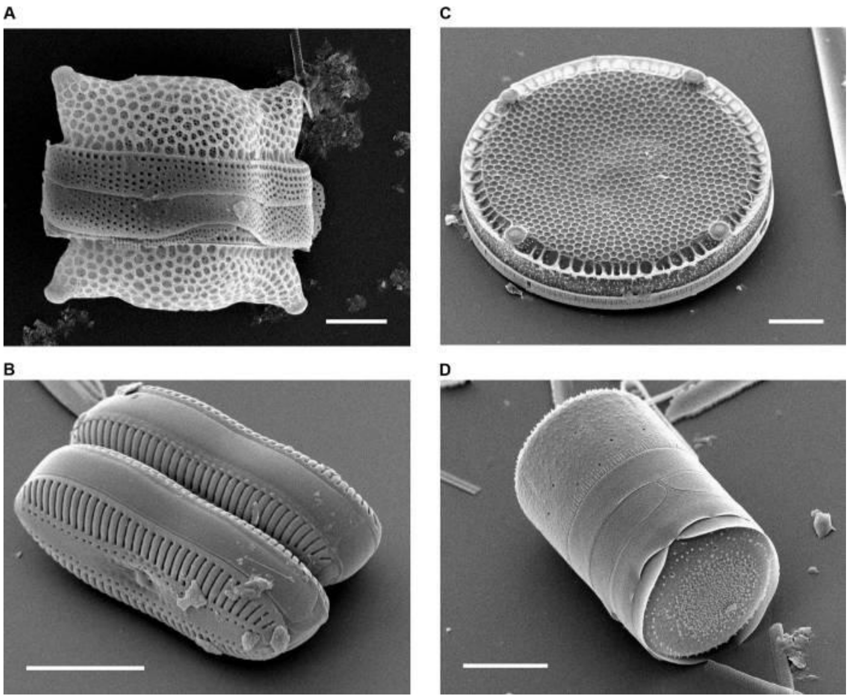


Figure 1.1: Electron microscope scan of four diatom frustules. A) *Biddulphia reticulata (size bar = 10 µm), B)Diploneis sp. (size bar = 10 µm), C) Eupodiscus radiatus (size bar = 20 µm), D. Melosira varians(size bar = 10 µm) (Bradbury 2004).*

The frustules consist of two identical halves, with one of the halves being slightly larger and overlapping the other half. Diatoms are divided into two main types: centric diatoms, which often have circular symmetry, and pennate diatoms, which often have bilaterally symmetry (Bradbury 2004). By creating silica cell walls from dissolved silicon in the sea, diatoms play a central role in biogenic cycling of silicon in the world`s oceans. Frustules from dead diatoms accumulate on the sea floor as large deposits of silica. The frustules contain specie-specific pore patterns of nano to micro-scale that are genetically controlled, although external factors can influence the density and pore size. The ability of diatoms to produce three-dimensional structures of silica has made them attractive models for nanotechnology (Armbrust 2009).

It is estimated that diatoms arose in the Triassic period, about 250 million years ago, by endosymbiotic events (Figure 1.2). In the so-called primary endosymbiotic event, a eukaryotic heterotroph engulfed a cyanobacterium, forming the photosynthetic plastids that are found in land plants, red and green algae. Thousands of genes were subsequently transferred from the endosymbiont to the host nucleus. A secondary endosymbiotic event occurred about 500 million years later, where a different heterotroph engulfed a red-algae. Genes were transferred from the endosymbiont red-algae nucleus and from the plastids genomes to the host nucleus. This transformation resulted in the formation of the Stramenopiles, which include diatoms, brown macroalgae and plant parasites. The emergence of diatoms resulted in a major shift in global organic carbon cycling, decreasing atmospheric CO_2 and increasing O_2 concentrations (Armbrust 2009).

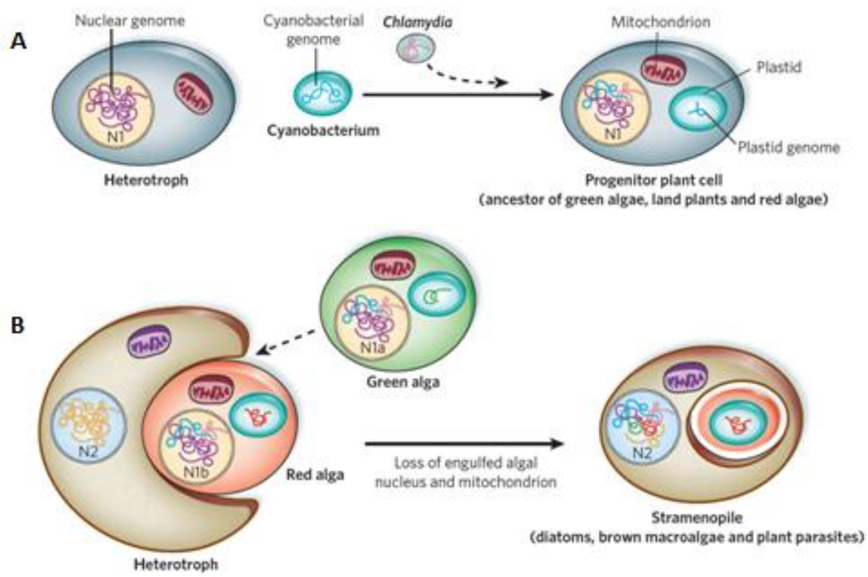


Figure 1.2: Origin of diatoms through endosymbiosis. A) Primary endosymbiosis. Heterotroph engulfs a cyanobacterium and integrates parts of the cyanobacterial genome into the host nucleus (N1). B) Secondary endosymbiosis. A serial engulfment by a heterotroph of a green and red algae (the former shown by a dotted arrow) results in the formation of the Stramenopile algae. The algal mitochondria and nucleus are lost, while important algal nuclear and plastid genes are integrated in the host nucleus (N2) (Armbrust 2009).

Despite the importance of algae in aqueous ecosystems and the carbon cycle, little is known about the carbohydrate pathways in these algae. The diatoms *Thalassiosira pseudonona* and *Phaedactylum tricornutum* have been fully sequenced and represent two of the major classes of diatoms: the bi/multipolar centrics (Mediophyceae) and the pennates (Bacillariophyceae) respectively. *P. tricornutum* and *T. pseudonana* share 57% of their genes, and 1,328 of these shared genes are absent from other sequenced eukaryotes (Bowler, Allen et al. 2008). The sequencing of *P. tricornutum* and *T. pseudonana* has made it easier to identify genes and gene families of interest in these diatoms. Methods for both transient and stable transformation of *P. tricornutum* and *T. pseudonana* have been established and have made molecular studies of these diatoms easier and more sophisticated (Falciatore and Bowler 2002; Poulsen Nicole 2006).

1.2 DNA repair

DNA damage may occur from many different agents, including radiation, spontaneous mutations and chemical mutagens. If the DNA is replicated before the damage is repaired, mutations may occur through incorporation of incorrect bases. DNA damage may also lead to cell death by creating chromosome breaks or by blocking replication. All organisms contain control systems to repair damaged DNA. These DNA repair systems are critical for protecting the genes against mutations and include several different approaches. Some act in a more general manner, by recognizing DNA structure abnormalities, while others are more specific, by recognizing specific chemical defects (Clark 2010). Defects in the DNA repair systems can lead to a hyper mutable phenotype and mutations on growth control genes. Two main principles for DNA-repair are known: Direct base repair by an enzyme without using the information of the complementary strand and base excision repair. During base excision repair, the damaged base is removed and exchanged by DNA repair synthesis. Base excision repair include: base excision repair (BER), mismatch-repair (MMR), nucleotide excision repair (NER) and recombination repair. BER is believed to be the most commonly used repair mechanism (Krokan and Slupphaug 1998).

1.3 Base excision repair (BER)

The BER pathway is divided into the sub-pathways: single-nucleotide BER (SN BER) and long-patch BER (LP BER). The pathways differ in the enzymes involved and the size of the repair patch. The first step in BER is recognition and removal of the damaged nitrogenous base by DNA glycosylases. SN BER involves removal and replacement of a single damaged nucleotide, while the LP BER pathway involves the removal and replacement of two or more nucleotides. Both pathways precede sequentially, one step after another (Figure 1.3). The repair process starts with the removal of the damaged base from the sugar phosphate backbone by enzymatic removal or by spontaneous chemical hydrolysis, creating an apurinic/apyrimidinic (AP) site while leaving the sugar-phosphate backbone intact. AP endonuclease 1 then recognizes and cleaves the deoxyribose at these sites. DNA polymerase β, a multifunctional deoxyribose phosphate lyase, and nucleotidtransferase enzyme, fill the gap created by AP endonuclease. Finally, DNA ligase 1 or XRCC1-DNA ligase III complex seals the nick (Krokan and Slupphaug 1998; Zhu 2009; Prasad, Shock et al. 2010).

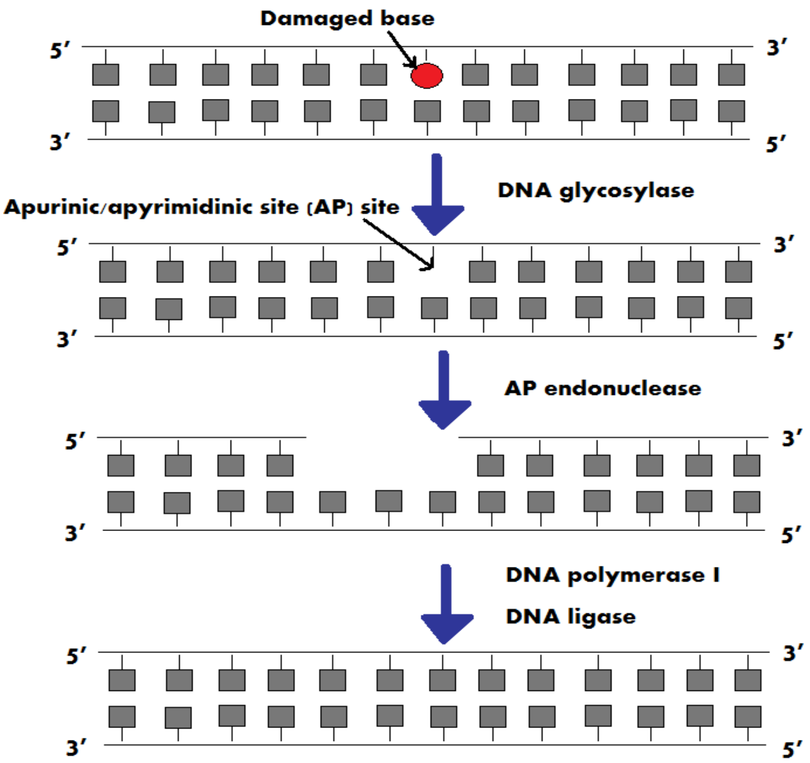


Figure 1.3: Single nucleotide base excision repair (SN BER). The damaged or inappropriate base is recognized and removed by DNA glycosylase, creating an AP site. AP endonuclease cleaves the sugar-phosphate backbone at the AP site. DNA polymerase 1 fills the gap and DNA ligase seals the nick. Adapted from: (Clark 2010)

Detection of a single damaged base in the genome seems like an impossible task, especially since BER lesions usually do not cause notable DNA distortions. Human cell base lesions happen at a rate of approximately 10^4 base lesions per cell per day, which is about one lesion per 10 s. How DNA glycosylases manage to recognize these lesions efficiently in a genome of about 14 billion nucleotides is not fully understood (Jacobs and Schar 2012).

DNA glycosylases are subdivided into monofunctional and bifunctional enzymes by their catalytic mechanism. The monofunctional glycosylases only perform base excision by performing a nucleophilic attack on the N-glycosidic bond with an activated water molecule. Bifunctional glycosylases form a Schiff-base intermediate by using an amino group of a lysine chain. After the Schiff-base intermediate is formed, the bifunctional glycosylase cleaves the DNA backbone 3' to the lesion (Jacobs and Schar 2012).

1.4 Uracil N-glycosylase (UNG)

Uracil may be misincorporated in DNA as dUMP during replication, resulting in U: A mismatches, or by spontaneous or enzymatic deamination of cytosine, resulting in U:G mismatches. A U:A mismatch is not miscoding, but may result in cytotoxic and mutagenic AP intermediates. Cytosine deamination has been calculated to occur at a rate of 100-500 per human cell per day. Uracil DNA glycosylases (UDG) are expressed in all living organisms to prevent the above lesions. UDG's remove uracil from deoxyribose and initiate the BER pathway (Kavli, Sundheim et al. 2002).

The first described UDG's (Family-1) are specified for efficient uracil excision from both single and double stranded DNA, regardless of second strand base partner. They are only specified for uracil in DNA and show negligible activity for cytosine or thymine in DNA bases and uracil in RNA. The specificity of family-1 UDG's is based on the structure of the active site, which binds uracil in a "pocket" formed by highly conserved residues (Figure 1.4). These residues provide a selective and specific uracil binding site, while disfavoring binding of other bases. A set of hydrogen binding interactions in the bottom of the "pocket" enables selection between cytosine and uracil, the two pyrimidine bases that are able to enter the active site (Pearl 2000).

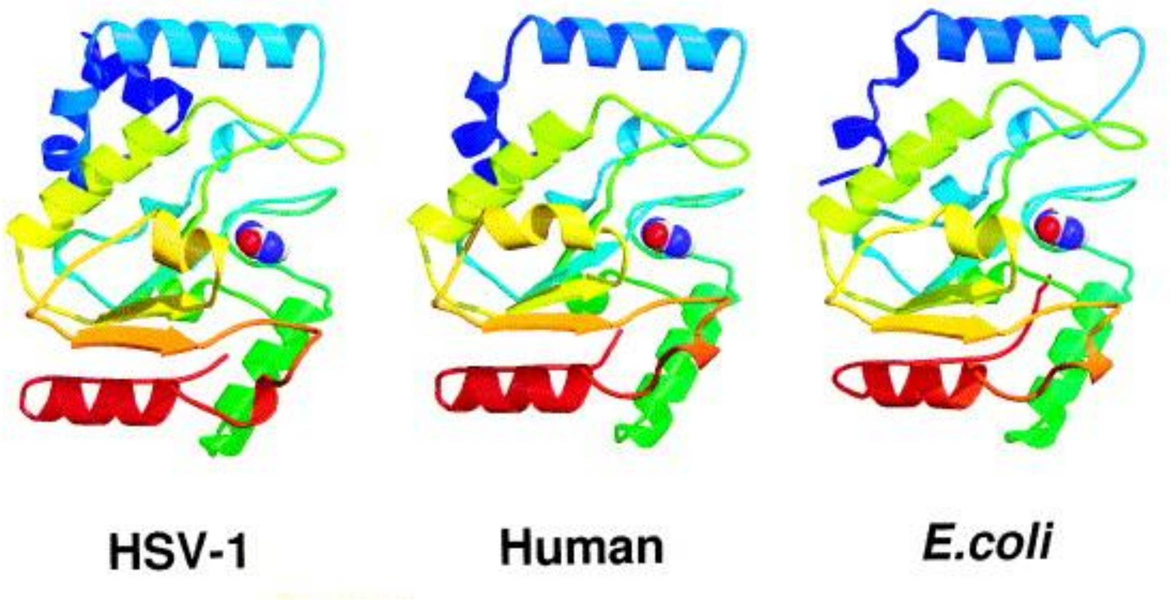


Figure 1.4: Family-1 Uracil-DNA glycosylases. Secondary crystal structure of herpes simplex virus 1 (1UDH), human (1SSP) and *E. coli* (2EUG) UDG's. The N-terminus is colored blue and C-terminus red. The bound uracil base is shown as a CPK model (Pearl 2000).

The first described UNG was identified in *E. coli,* and UNG homologs were later characterized in many other organisms. E. coli and yeast with base mutations in the UNG gene have shown an increase in spontaneous mutation rate. Depending on the mutated base, the spontaneous mutation rate increases from 5- to 30-fold (Slupphaug, Eftedal et al. 1995).

Five different UDGs have been identified in mammals: nuclear UNG1, mitochondrial UNG2 and the nuclear proteins SMUG1, TDG and MBD4 (Visnes, Doseth et al. 2009). These glycosylases cleave uracil residues from the DNA and share a common fold and overall architecture, despite their low sequence homology (≈8%). All five UDGs repair additional lesions and have different substrate and context preferences. The exact mechanism of how these glycosylases selectively remove the target lesion is not well known (Darwanto, Theruvathu et al. 2009).

UNG1 and UNG2 are encoded by alternative promoter usage and mRNA splicing of the same UNG genes. They appear to constitute the quantitatively dominant UDG determined by activity assays with U:A substrates in human cell-free extracts. UNG1 is composed of 304 amino acids, while UNG2 is composed of 313 amino acids. The first 35 N-terminal amino acids of UNG1 and the first 44 N-terminal amino acids of UNG2 are required for their mitochondrial and nuclear localization, respectively (Chen, Mosbaugh et al. 2004). These two glycosylases belong to the highly conserved UNG family, which is present in a large number of eukaryotes, bacteria and large eukaryotic DNA viruses (Kavli, Sundheim et al. 2002). UNG family members are highly conserved at both the amino-acid sequence and gene structure levels. Human and yeast UNG share 40.3% amino acid sequence similarity, and exon-intron boundaries are identical between human, mouse and fish genes. This indicates 450 million years of unchanged exon-intron organization of UNG genes. UNG2 is transported to the cell nucleus by interacting with the nuclear proteins proliferating cell nuclear antigen (PCNA) and replication protein A (RPA). Its main function is to prevent uracil from getting incorporated in the DNA. Mice cells lacking UNG accumulate uracil in their DNA at an approximately 100-fold increased rate. Despite the high increase of uracil into their DNA, they did not show a mutation phenotype. The mice do however develop B-cell lymphomas and show changes in their antibody diversification. A subgroup of hyper-igM syndrome patients, impaired class switch recombination, has also been associated with human UNG gene mutations (Jacobs and Schar 2012). UNG2 is most likely the only UDG participating as a part of the mechanisms in B-cell ig somatic hypermutation and class switch recombination (Visnes, Doseth et al. 2009).

All UNGs examined have shown activity towards uracil in both ssDNA and dsDNA, but have a preference for uracil in ssDNA. UNG preferences for uracil are in order of ssU > U:G >U:A (Slupphaug, Eftedal et al. 1995). Despite being selective for uracil, UNG also shows activity towards closely related cytosine lesions that are created by oxidative stress or gamma irradiation, but at a lower rate compared to uracil (Krokan, Standal et al. 1997). Cytosine deamination occurs at an approximately 300-fold faster rate in ssDNA compared to dsDNA (Lindahl 1993). Deamination of ssDNA in replication forks can lead to mutations if left unrepaired. UNG preference for ssDNA could therefore be of critical biological importance for maintaining the DNA (Hagen, Kavli et al. 2008).

Uracil recognized by UNG is tightly fitted between amino-acid residues at the bottom of the UNG pocket. These amino-acid residues position the N-glycosidic bond for hydrolysis. A strategically positioned activated water molecule attack and cleaves the N-glycosidic bond, creating a free base or AP-site (Figure 1.5).

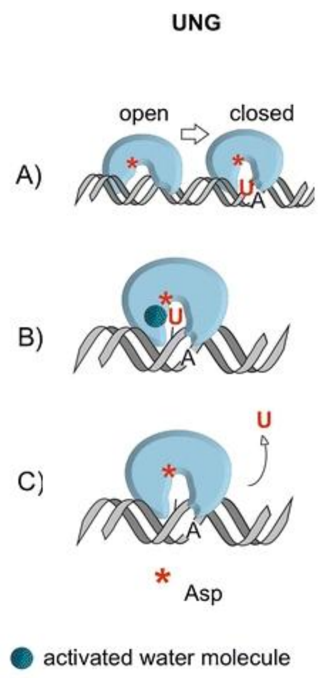


Figure 1.5: Mechanism of base removal by the monofunctional glycosylase UNG. A) Uracil is recognized and captured by UNG. B) Uracil is flipped out of the DNA double helix and into the catalytic pocket of UNG C) Nucleophilic attack by an activated water molecule. Red asterisk: conserved aspartic acid residue, which is crucial for positioning of the activated water molecule (Jacobs and Schar 2012).

The catalytic mechanism of the N-glycosidic cleavage has been established by structural and mutational analysis. A conserved histidine polarizes the N-glycosidic bond, making it a target for nucleophilic attack. The activated water molecule is positioned and deprotonated by a conserved aspartate. C1 of the deoxyribose is attacked by the deprotonated water molecule, which cleaves uracil and results in an AP-site (Jacobs and Schar 2012).

Studies of the kinetic reaction mechanisms of UNG have led to a proposed ”pinch-pull-push” catalytic mechanism (Figure 1.6). The mechanism was postulated as a sequence of three events. First the UDG scans the minor groove of the DNA for uracil residues. It uses the 4 Pro-Ser loop and the Gly-Ser loop to bind, bend and compress the dual DNA backbone. Second, when uracil is located in the DNA, the uracil residue is flipped out of the double helix and replaced by a conserved leucine side chain (Leu 272). Third, the uracil base is captured and stabilized in the UDG active pocket, where the N-glycosidic bond is cleaved (Chen, Mosbaugh et al. 2004).

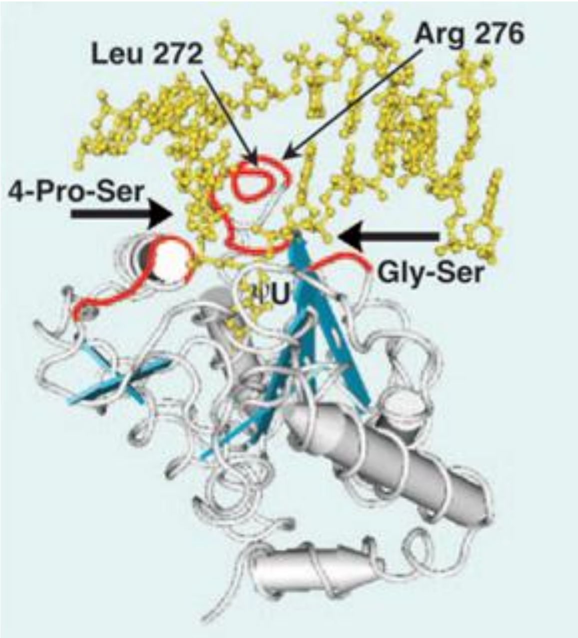


Figure 1.6: Drawing of tertiary structure of human UNG bound to DNA. DNA is shown in yellow and the view is set to be looking into the major groove. ψU: flipped out uracil residue, silver cylinders: α-helices, blue strands: β –Sheets, red tubes: three amino acid sequences critical for proposed catalytic pinch-pull-push mechanism. The 4 Pro-Ser and Gly-Ser loop pushes the deoxyribose phosphate backbone together, while the Leu 272 loop, containing Arg 276, penetrates into the DNA double helix and occupies the flipped out uracil residue. Uracil is captured and stabilized by conserved amino acid residues in the UNG pocket (Chen, Mosbaugh et al. 2004).

1.5 NEI-type glycosylases (NEIL)

Reactive oxygen species (ROS), such as O_2, H_2O_2, and OH- radicals, are byproducts of respiration and important genotoxic agents. ROS are also produced during inflammatory responses and are believed to be causes of many pathophysiologies and aging. The pathophysiologies range from cancer, cardiovascular diseases, Alzheimer's disease to rheumatoid arthritis. The genotoxicity of ROS compounds is based on their ability to react with DNA and create different DNA base lesions and strand breaks. These base lesions are primarily repaired by the BER pathway in both eukaryotes and bacteria. Different DNA glycosylase AP lyases excise the damaged base and initiate the BER pathway (Hazra, Izumi et al. 2002).

Based on the presence of certain structural motifs these glycosylases are divided into the HhH superfamily and the Fpg/Nei family. The glycosylases are divided further based on their substrate specificity. The Fpg/Nei family enzymes share a two-domain structure consisting of an N-terminal β-sandwich with two α-helices and a C-terminal domain of four α-helices. Two of the C-terminal helices are important for a helix-two-turn-helix motif and for the zinc finger motif β-strands (Figure 1.7). The first family member of the Fpg/Nei family to be identified was the *E. coli* formamidopyrimidine DNA glycosylase (Fpg). Fpg was recognized by its ability to remove methyl-formamidopyrimidines. 8-oxoG was later found to be the main substrate of Fpg (Imamura, Averill et al. 2012). 8-oxoG derives from oxidized guanine and can lead to G:C to A:T mispairing during replication if left unrepaired (Krokan, Standal et al. 1997). Structural studies of Fpg with 8-oxoG substrates have revealed insights on the lesion presentation of the active site of Fpg. The αF-β9/10 loop recognizes and stabilizes the everted 8-oxoG. It has been hypothesized that the 8-oxoG recognition occurs before it is flipped out of the DNA helix (Imamura, Averill et al. 2012).

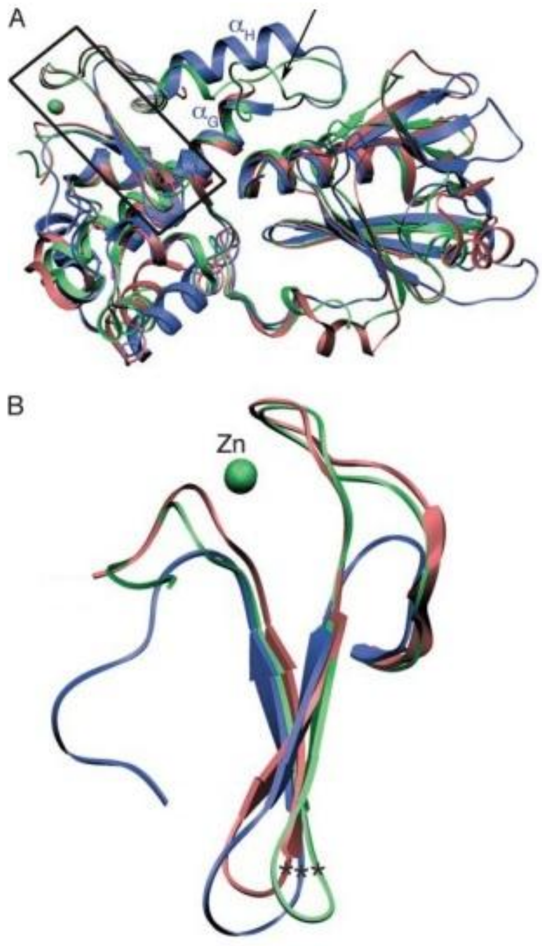


Figure 1.7: Human NEIL1 structure comparison. A) Human NEIL1 are shown in blue, EcoNei in pink and TthFpg in green. The boxed region is surrounding the zinc-finger motif. The arrow shows the αF-β9/10 loop. B) A close-up comparison of the zinc-finger motif. The same color codes as in A are used. The position of the Cα conserved arginine is marked with asterisks (Doublie, Bandaru et al. 2004).

The Nei family glycosylases are bifunctional enzymes, with both glycosylase and lyase activity. They recognize and cleave oxidized pyrimidine, but have also been shown to cleave further oxidation products of 8-OxoG, guanidinohydantoin and spiroinodihydantoin. Human NEIL1 and Mimivirus Nei1 (MvNei1) are orthologue Nei family glycosylases. By testing commercial available substrates with MvNei1, the enzyme has been shown to recognize and remove thymine glycol (Tg) and 5-hydroxyuracil (5-OHU), but shows negligible activity with 8-oxoG. Deamination of cytosine generates 5-OHU and oxidation of thymine generates Tg. 5-OHU can be in both anti and syn conformation, while Tg is only in the anti conformation in the enzyme's active site. The crystal structure of MvNei1 has been identified and is the first Nei glycosylase complex with a damaged base to be presented (Figure1. 8).

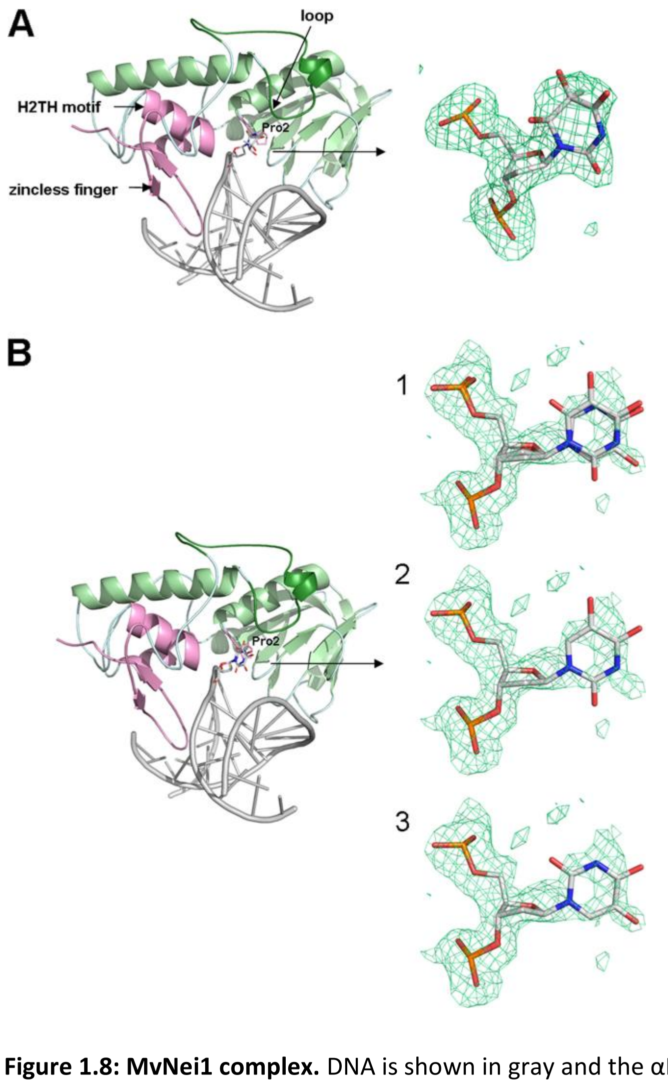


Figure 1.8: MvNei1 complex. DNA is shown in gray and the αF-β9/10 loop in dark green. The helix-two-turn-helix (H2TH), the zinc-less finger and catalytic proline are shown in pink. A) MvNei1 in complex with Tg. B) MvNei1 in complex with 5-OHU in syn and anti conformation (1), anti conformation (2) and syn conformation (3) (Imamura, Averill et al. 2012).

The structure is very similar to hNEI1, EcoNei and other Fpg/Nei glycosylase structures. Figure 1.8 shows that MvNei1 has few interactions with the damaged base, and suggests that base recognition happen before the base flipping occurs (Imamura, Averill et al. 2012).

Homo sapiens contains three homologs of Nei; Nei-like (NEIL)1, NEIL2 and NEIL3. The homologs have been cloned and NEIL1 and NEIL2 have been expressed and characterized at some degree. NEIL1 activity apparently serves as an Nth backup mechanism, as seen in nullizygous Nth mice. NEIL1 knock-down stem cell lines have been shown to become hypersensitive to ionizing radiation, suggesting that NEIL1 is essential in BER in mammalian cells. NEIL1 and NEIL2 differ in substrate preference, but both recognize oxidized pyrimidines and form a Schiff base with their substrates. NEIL2 and NEIL3 contain a zinc finger motif, while NEIL2 contains a motif that strongly resembles the zinc finger motif. All three homologs share the DNA-binding helix-two-turn-helix motif. NEIL1 and NEIL2 share a catalytic N-terminal proline, which in NEIL3 is replaced by a valine (Doublie, Bandaru et al. 2004).

NEIL1 is best characterized of the three NEIL homologs. NEIL1 activity has not only been shown towards oxidized pyrimidines, but also against several mutagenic and cytotoxic DNA lesions. Substrates of NEIL1 include the ring-opened formamidopyrimidines (FapyA and FapyG), which are among the major cytotoxic lesions of oxidative DNA damage. Other DNA lesions NEIL1 have been reported to remove include two *cis*-Tg stereoisomers, 5,6-dihydrouracil and two major secondary byproducts of 8-oxoG, guanidinohydantoin and spiroiminodihydantoin (Lindahl 1993; Zhao, Krishnamurthy et al. 2010). NEIL1 expression and protein levels are up-regulated during S-phase, and NEIL1 has a preference for ssDNA. The enzyme also interacts with replication-associated proteins, suggesting that NEIL1 have a role in the replicative process (Hazra, Izumi et al. 2002; Theriot, Hegde et al. 2010). NEIL1 interaction studies show that NEIL1 can initiate either the SN-BER or LP-BER pathway (Hegde, Theriot et al. 2008).

1.6 Dual DNA glycosylase (DDG) in diatoms

UNG and NEIL have previously only been identified as single domain proteins (Figure 1.9A). Sequence studies of the diatoms *Phaeodactylum tricornutum* (Pt) and *Thalassiosira pseudonana* (Tp) have revealed that both diatoms encode a protein that contains two glycosylase domains. The glycosylase has been named “Dual DNA glycosylase” and has only been discovered in diatom genomes. DDG contain an N-terminal NEIL domain and a C-terminal UNG domain, connected with a linker region (Figure 1.9B).

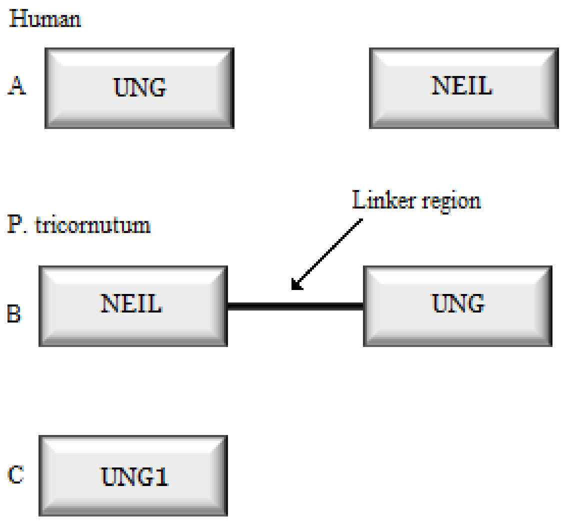


Figure 1.9: Domain comparison of hUNG, hNEIL, PtDDG and PtUNG1. A) hUNG and hNEIL domain organization. Both glycosylases are single domain glycosylases with an N-terminal signal sequence. B) PtDDG domain organization. PtDDG contains an N-terminal signal sequence, an N-terminal NEIL domain and a C-terminal UNG domain. The UNG and NEIL domains are connected with a linker region. C) PtUNG1 domain organization. PtUNG1 is a single domain glycosylase with an N-terminal signal sequence. Modified from: (Hem 2010)

DDG has later been found in the pennate diatoms *Fragilariopsis cylindrus* and *Seminavis robusta,* suggesting that DDG is a diatom specific protein. PtDDG has the following amino acid (AA) organization: AA 1-36: N-terminal signal region, AA 37-340: NEIL domain, AA 341-380: linker region and AA 381-598: UNG domain. Both *P. tricornutum* and *T. pseudonana* encode an UNG homolog in addition to the DDG UNG domain. The UNG homolog (UNG1) has a "normal" domain organization, containing an N-terminal signal sequence and a UNG domain, as seen in other eukaryote organisms (Figure 1.9C) (Hem 2010). Sequence comparison of *P.tricornutum* and *T. pseudonana* DDG UNG and NEIL domains with homologous proteins in other organisms shows that important structural and functional motives have been preserved in both domains. The NEIL H2TH and the zinc finger motif are conserved in the DDG NEIL domain. The DDG NEIL domain also contain the two catalytic N-terminal Pro and Lys residues found in the active sites of other NEIL proteins. The DDG UNG domain contains at least five conserved UNG motifs: the water-activating motif, the proline-rich motif, the glycine-serine motif, the uracil-recognition motif and the intercalation loop motif. These five UNG motifs have also been found to be conserved in *P.tricornutum* and *T. pseudonana* UNG1 (Hem 2010).

1.7 Aim of study

The overall aim of the study is to gain further knowledge about the Dual DNA Glycosylase from *P. tricornutum*. A protein with two DNA glycosylase domains is unique for diatoms, and has never been characterized. The main aim of this study is to clone, express and purify sufficient amounts of PtDDG, as full-length DDG, N-terminal NEIL domain and C-terminal UNG domain for further downstream analysis. It is of interest to use the purified DDG for activity assays to be able to find optimum conditions and specific activity of the enzyme. A part of the study also aims to visualize the intracellular localization of PtUNG1 and PtDDG full length. This part is done by transformation (biolistic bombardment) of a plasmid containing recombinant PtDDG and PtUNG1 with yellow fluorescent protein (YFP) attached to the final product.

2. Methods

In this section the main methods used during the master project are described. Each method starts with a short principal description. Media, buffers, primers, vectors and calculations used are listed in the appendices.

2.1 DNA work

2.1.1 Polymerase chain reaction (PCR)

PCR allow very small amounts of DNA molecules to be amplified into enough DNA for further downstream analyses. The reaction uses a pre-existing DNA sequence as template and DNA polymerase for amplification by repeated cycles of replication and strand separation. Two primers are needed to initiate the PCR synthesis. Primers are small single-stranded DNA oligonucleotides that match the sequence of either end of the target DNA (Clark 2010).

Method

DDG domains were amplified by PCR as: full-length DDG, UNG domain and NEIL domain. Plasmids produced after running the QuickChange® II XL Site-Direct Mutagenesis were used as template. The following PCR mixture were mixed in 0,2 ml PCR tubes:

Autoclaved MilliQ water	35,75	µl
10 x Ex Taq Buffer	5	µl
dNTP mix	4	µl
Primer 1	2	µl
Primer 2	2	µl
Template	1	µl
Ex-Taq polymerase	0,25	µl
Total volume	50	µl

The primers used are listed in appendix 3. Each primer introduced a new restriction site (NcoI/EcoRI) to the PCR product for use in downstream manipulation. The tubes were briefly centrifuged on a tabletop centrifuge before the PCR run.

The PCR block was preheated to 105 °C. The following PCR parameters were used:

Temperature °C	Time
94 °C 50 °C 68 °C	30 sec. 60 sec. 2 min.
Cycles	4
94 °C 54 °C 68 °C	30 sec. 60 sec. 2 min.
Cycles	20

The annealing temperature was kept at 50 °C for four cycles to ensure primer binding. The temperature was increased to 54 °C for the next 20 cycles to prevent non-specific primer binding. The PCR samples were stored at – 20 °C (Hem 2010).

2.1.2 QuickChange mutagenesis

The method uses individually designed mutagenic primers and DNA polymerase to synthesize a mutant strand. The methylated and hemimethylated parental strands are digested by DpnI. The mutated strand is transformed into competent cells for nick repair and propagation (Stratagene).

Method

The method was performed after Stratagene: QuickChange® II XL Site-Direct Mutagenesis kit protocol. The PCR cycling parameters were set according to Stratagene: QuickChange® II XL Site-Direct Mutagenesis kit protocol. The extension step was set to 5 minutes. The primers used are listed in the appendix. The template strand used (PtDDG donor) was provided by earlier master student Cecilie Anne Dahl Hem (Hem 2010).

2.1.3 DNA sequencing: dideoxy sequencing

By using a mixture of deoxynucleoside triphosphate and fluorophore labeled dideoxynucleotides (one color for each of the four bases) during PCR, the DNA synthesis is terminated at random sites. This crates sub-fragments of all possible lengths of the DNA template. The sub-fragments are separated by size through capillary electrophoresis. The template sequence is then read by passing the separated fragments past a detector that recognizes the different fluorophores (Clark 2010).

Method

The sequencing reaction was performed by using BigDye® Terminator v3.1 Cycle Sequencing Kit. The method was performed according to the kit specifications.

The samples were stored at -20°C before they were shipped to the University of Tromsø (UiT). Samples were sequenced by technical staff at UiT.

2.1.4 DNA separation: Agarose gel electrophoresis

Agarose gel electrophoresis is a method used for separating nucleic acid molecules. By adding the molecules to an agarose gel and passing an electric current through the gel, the molecules are separated by size (Clark 2010).

Method

1. 50 ml TAE buffer (Appendix 1) was mixed with 0,5 g agarose (1% agarose gel).
2. The solution was heated to its boiling point in a microwave oven.
3. The solution was allowed to cool for 3-4 minutes on a shaker.
4. 2,5 µl Gel Red ™ was added to the solution.
5. The solution was allowed to cool for 10 minutes on a shaker before it was poured into Bio-Rad Sub-Cell® GT agarose gel electrophoresis system. The gel system was put together according to Bio-Rad Sub-Cell® GT agarose gel electrophoresis system manual. The gel was allowed to solidify.
6. Samples were mixed with Fermentas 6X Loading Dye, vortexed and briefly centrifuged in a tabletop centrifuge.

7. The gel was submerged in 1X TAE running buffer before the samples were added. 6 µl Fermentas GeneRuler ™ 1kb DNA ladder were used as molecular marker.
8. The gel was run at 80 V for 50-60 minutes.
9. DNA was visualized by Bio-Rad GelDoc 2000.

2.1.5 DNA purification from agarose gel

DNA can be purified by passing the DNA containing sample through a DNA binding column, such as a silica column. The silica column binds nucleic acids at low pH and high salt concentration. To release the nucleic acid from the column, the pH is increased and the salt concentration is decreased (Clark 2010).

Method

The purification was performed by using Wizard® SV Gel and PCR Clean-Up System kit. The method was performed according to the kits specifications. The purified DNA was eluted with 30 µl Nuclease-Free Water, instead of 50 µl Nuclease-Free Water, as recommended in the Wizard® SV Gel and PCR Clean-Up System protocol.

2.1.6 DNA quantification

Because of the aromatic rings in DNA and RNA bases, their absorption maximum is at 260 nm. By irradiating a DNA or RNA solution with 260 nm UV light, the amount of absorbed light is proportional with the amount of DNA or RNA in the sample. By measuring and calibrating a calibration curve with samples of known DNA or RNA concentrations, the amount of DNA or RNA in an unknown sample can be determined (Clark 2010).

Method

The DNA quantification was performed using a NanoDrop™ 1000 spectrophotometer. The method was performed according to the NanoDrop™ 1000 spectrophotometer manual. 1 µl MilliQ water was used as blank.

2.1.7 Restriction enzymes

Restriction enzymes are endonucleases that cut dsDNA at specific base sequences called recognition sites. Restriction enzymes are commonly used for cutting out a gene of interest and relocating the gene into a vector for further manipulation (Clark 2010).

Method

1. Reagents were kept on ice and mixed in a 0,2 µl PCR tube.

10 x Buffer	2 µl
Restriction enzyme 1	0,5 µl
Restriction enzyme 2	0,5 µl
DNA	1-2 mg
* 10 x BSA	2 µl
Autocleaved MilliQ water	To final volume
Total volume	20 µl

* = Only added if recommended by the restriction enzyme manufacturer.

2. The samples were briefly centrifuged in a tabletop centrifuge and incubated for two hours at 37 °C.
3. The samples were stored at – 20 °C.
4. Restriction cleavage was verified by gel electrophoresis (Hem 2010).

2.1.8 DNA ligation: plasmid vectors

DNA ligase is an enzyme that seals the nicks in the phosphodiester backbone of DNA (Alberts 2008).

Method

1:3 molar ratio of insert: vector was used in the ligation reactions. The amount of insert and vector were calculated for each mixture.

1. Reagents were kept on ice and mixed in a 0,2 µl PCR tube.

Vector	Calculated
Insert	Calculated
Ligase 10 x Buffer	1-2 µl
Autocleaved MilliQ water	To final volume
Total volume	10-20 µl

2. The samples were briefly centrifuged in a tabletop centrifuge and incubated at 16 °C overnight.
3. The samples were stored at – 20 °C (Hem 2010).

2.1.9 Heat-shock transformation

Heat-shock transformation is a method to facilitate the entry of foreign DNA into bacterial cells. Competent cells are incubated on ice prior to a short heat-shock (37-45 °C). The heat-shock destabilizes the cell walls temporarily; the foreign DNA is able to enter some of the cells. Nutrient medium is then added to the cells, so they can grow and express the phenotype of the foreign DNA (e.g. antibiotic resistance). The cells can be plated on selective medium, where only the transformed cells will be able to survive (Reece 2009).

Method

1. 100 µl RbCl competent *E. coli* cells were thawed on ice.
2. 2-10 µl DNA was added and mixed gently. The samples were incubated on ice for 30 minutes.
3. Cells were heat-shocked in a 42 °C water bath for 45-55 second.
4. The cells were incubated on ice for 2 minutes.
5. 1 ml preheated (37 °C) LB medium (Appendix 1) was added.
6. The samples were incubated for 1 hour at 37 °C and 220 rpm in a shaking incubator.
7. 200 µl of the sample were plated on antibiotic selective LB plates (Appendix 1).
8. Plates were incubated at 37 °C overnight.
9. Colonies were picked for further molecular manipulation. The plates were sealed with parafilm and stored at 4 °C (Hem 2010).

2.1.10 Biolistic transformation of diatoms

Biolistic transformation is based on DNA delivery via a biolistic particle delivery system. The system fires high velocity metal particles coated with DNA into the target cells. This allows foreign genes to be introduced and expressed in the cells (Bio-Rad ; Kroth 2007).

Method

Biolistic transformation of diatoms was performed by using Bio-Rad Biolistic® PDS-1000/He system (Figure 2.1). The system uses high-pressure helium and partial vacuum to shoot a macrocarrier loaded with microscopic tungsten or gold microcarriers toward the target cells. The tungsten or gold microcarriers are coated with the material for transformation. The macrocarrier is stopped by a stopping screen, but the coated microcarriers continue towards the target cell to penetrate and transform them (Bio-Rad).

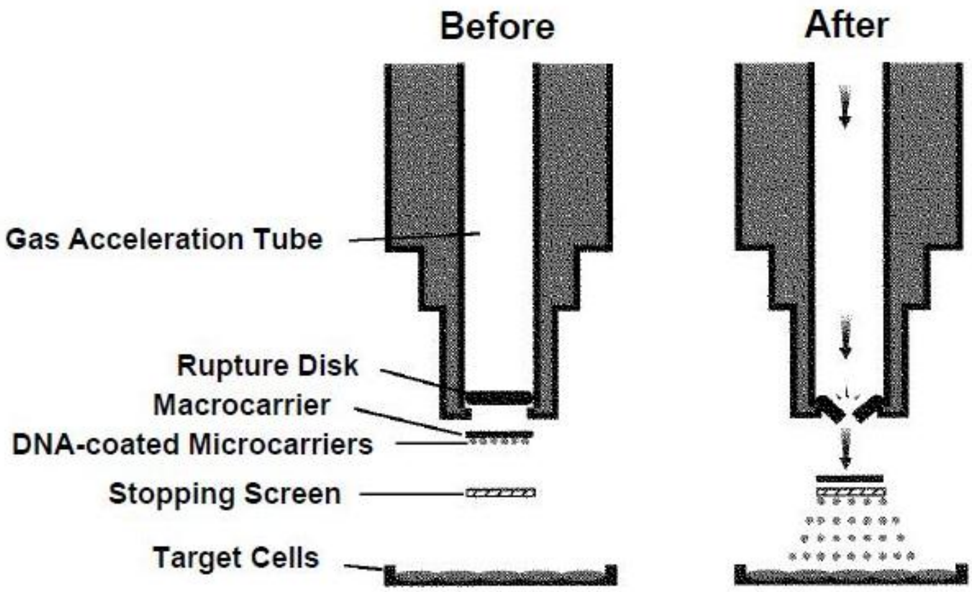


Figure 2.1: Schematic overview of Bio-Rad Biolistic® PDS-1000/He system. The system uses helium to create enough pressure in the gas acceleration tube to break the rupture disk. The released pressure shoots the macrocarrier holding DNA-coated microcarriers. The macrocarrier is stopped when reaching the stopping screen, but the DNA-coated microcarriers continue towards the target cells. The DNA-coated microcarriers penetrate and transform the target cell. The launch velocity is dependent on the helium pressure (rupture disk used), vacuum amount in the bombardment chamber and the distances between rupture disk to macrocarrier, macrocarrier to stopping screen and stopping screen to the target cells (Bio-Rad).

Preparation of diatoms prior bombardment

1. An exponentially growing *P. tricornutum* culture was centrifuged for 10 minutes, 3000 rpm at room temperature. The pellet was resuspended in a small amount of f/2 medium (Appendix 1).
2. Diatom concentration was found by using a Bürker counting chamber. Cell density was adjusted to approximately 5 x 10^7 cells per 500 µl solution.
3. 500 µl of the adjusted diatom solution was plated onto f/2 agar (Appendix 1) plates without antibiotic selection. The plates were allowed to dry in a laminar flow hood before they were incubated in a climate controlled growth room. Growth room specification: 100 µmol photons $m^{-2} s^{-1}$ in a 12 hour light: 12 hour dark photoperiod at 20 °C (Hem 2010).

Preparation of tungsten particle for biolistic bombardment

Tungsten deagglomeration

60 mg tungsten particles (Bio-Rad: Tungsten M-17, 1,1 µm diameter) were mixed with 1ml 100% ethanol in a 1,5 ml eppendorf tube. The mixture was vortexed at full speed for 3 x 2 minutes (Hem 2010).

Washing the microcarriers

The eppendorf tubes containing the deagglomerated tungsten particles were centrifuged at 10 000 rpm for 1 minute at room temperature. The supernatant was removed and 1 ml autoclaved MilliQ water was added. Tungsten particles were resuspended by full speed vortexing. This step was repeated two times. 50 µl aliquots were made, while the eppendorf tube was kept under constant vortexing. Aliquots were stored at – 20 °C (Hem 2010).

DNA coating of the microcarriers

1. An aliquot of tungsten particles was thawed and resuspended by vortexing.
2. The aliquot was kept under constant vortexing while adding: 5 µg plasmid DNA, 50 µl 2,5M $CaCl_2$ and 20 µl 0,1M spermidine.

3. The mixture was vortexed for 2 minutes and kept on ice for 1 minute for the particles to sediment.
4. Particles were briefly centrifuged at 1000 rpm and the supernatant removed by pipetting.
5. The pellet was gently mixed with 140 µl 70% (v/v) ethanol and allowed to sediment for 3 minutes. The supernatant was removed.
6. The pellet was gently mixed with 140 µl 100% (v/v) ethanol and allowed to sediment for 3 minutes. The supernatant was removed.
7. The pellet was gently mixed with 50 µl 100% ethanol and kept on ice prior bombardment (Hem 2010).

Preparation of biolistic device and samples prior bombardment

1. The helium supply and vacuum pump was connected according to the Biolistic® PDS-1000/He system manual.
2. The chamber of Bio-Rad Biolistic® PDS-1000/He system and the different parts inside the system was wiped with 70% ethanol.
3. Rupture disks (1550 psi), macrocarriers, stopping screen and rupture disk retaining cap were sterilized in 50 ml sterile tubes filled with 70% ethanol and air dried.
4. 10 µl of particle solution containing the DNA coated microcarriers was pipetted onto the macrocarrier, while the eppendorf tube with the particle solution was kept under constant vortexing. The macrocarrier was allowed to dry.
5. The biolistic device was assembled according to the Bio-Rad Biolistic® PDS-1000/He system manual (Hem 2010).

Bombarding diatoms with the Bio-Rad Biolistic® PDS-1000/He system

1. The helium tank pressure regulator was set to 200 psi above the rupture disk tolerance.
2. Agar plates with diatoms were positioned at level 2 in the chamber (6 cm under the stopping screen).
3. The vacuum pump was used to evacuate the chamber to a vacuum pressure of 27,5 Hg.

4. The acceleration tube was filled with helium until the rupture disk ruptured.
5. The bombardment was repeated 5 times per agar plate. The bombardment was carried out at different regions of the agar plate.
6. After bombardment the agar plates were sealed with parafilm and incubated in a climate controlled growth room (described in: preparation of diatoms prior bombardment) for 72 hours (Hem 2010).

Transfer of bombarded diatoms to selective medium

The work was carried out in a laminar flow cabinet.

1. 1 ml f/2 medium containing vitamin solution and 100 µg/ml Zeocin was transferred to the bombarded agar plate. Cells were dispersed with a sterile glass rod.
2. The solution was divided onto two antibiotic selective (100 µg/ml Zeocin) f/2 agar plates and plated with a sterile plastic loop.
3. 0,5 ml f/2 medium containing 100 µg/ml Zeocin was pipetted onto the same agar plate. Remaining cells were dispersed with a sterile glass rod and transferred to a third selective f/2 agar plate. The solution was plated with a sterile plastic loop.
4. f/2 agar plates were allowed to dry, and were sealed with parafilm and placed in a climate controlled growth room (described in: preparation of diatoms prior bombardment) for 2-3 weeks (Hem 2010).

Picking transformed diatom colonies

1. Transformed colonies, seen as brown colonies without sharp edges, were examined with an epifluorescence stereomicroscope (Nikon) to check for yellow fluorescent protein expression (YFP).
2. Colonies were picked by a sterile plastic loop and transferred to a BD Falcon tissue culture flask (70 ml) containing 30 ml f/2 medium and 100 µg/ml Zeocin.
3. Tissue flasks were incubated in a climate controlled growth room (described in: preparation of diatoms prior bombardment) for 2-4 weeks.
4. Diatoms from incubated tissue flasks were visually inspected to check for expressed YFP with a confocal microscope (Hem 2010).

2.1.11 Plasmid isolation: Mini-prep

The method is basically divided into three steps:

1. The first step is alkaline lysis of the bacterial cell. Under these conditions chromosomal DNA become denatured, while the plasmid DNA remain intact in the solution.
2. The sample is then neutralized. Under these conditions the chromosomal DNA, proteins and high-molecular-weight RNA form an insoluble network, while the plasmid DNA still remain in the solution.
3. The insoluble network of chromosomal DNA, proteins and high-molecular-weight RNA are removed by centrifugation. The plasmid DNA will remain in the supernatant and can be further purified (Reece 2009).

Method

Plasmid DNA isolation was performed using the QIAGEN QIAprep® miniprep kit, according to the QIAprep® miniprep handbook.

2.1.12 Cloning and designing protein expression constructs

Molecular cloning means cutting out a gene of interest and relocating it into a vector for further manipulation (Clark 2010). TA cloning takes advantage of the terminal transferase activity of *Taq* polymerases, which adds a deoxyadenosine (A) to the 3' end of the PCR product. The TA cloning vector has an overhanging 3' deoxythymidine (T) residue, which allows PCR products generated by *Taq* polymerases to ligate efficiently into the vector (Invitrogen).

Method

PtDDG was cloned as Full-length (FL) DDG, C-terminal UNG domain and N-terminal NEIL domain into the pBAD-M30 expression vector. PBAD-M30 vector contains both a glutathione s-transferase tag (GST-tag) and a histidine tag (HIS-tag) for affinity purification (Figure 2.2). The vector also contains a Tobacco etch virus (TEV) protease site, which allows removal of the HIS-tag and the GST-tag after or during purification.

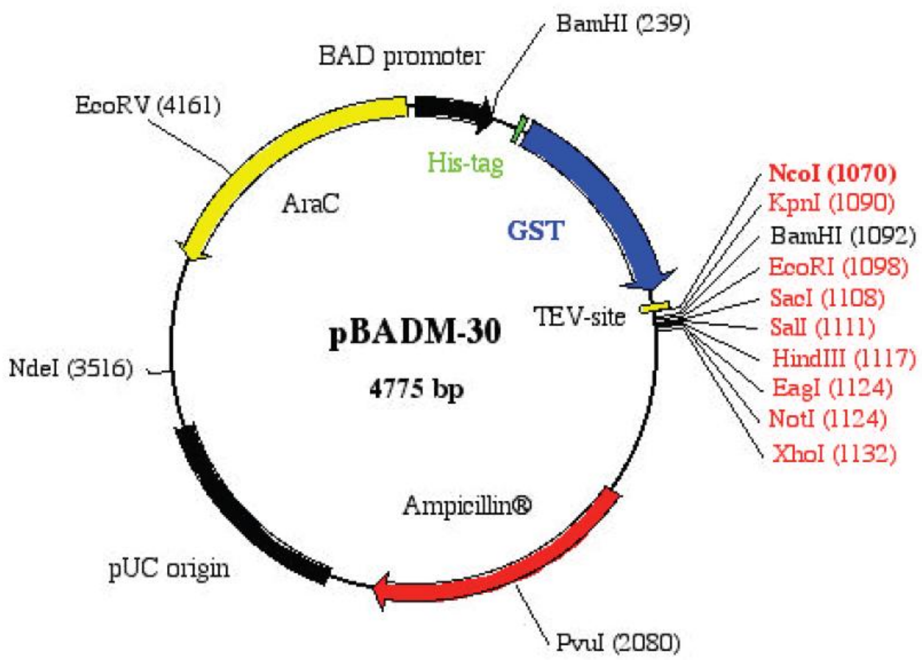


Figure 2.2: pBAD-M30 vector map. The map shows the key features of the pBAD-M30 expression vector.

TA Cloning

To obtain sufficient amount DNA for further downstream cloning, the DDG PCR products FL DDG, NEIL domain and UNG domain (see 2.1.1 for details) were subcloned into a TA cloning vector using TOPO® TA Cloning® Kit (Invitrogen). The method was performed according to the TOPO® TA Cloning® Kit manual. After the TOPO® cloning reaction was made, the samples were incubated at 30 minutes at room temperature.

The newly cloned TA vectors was transformed into RbCl competent *E. coli* cells and plated onto kanamycin LB plates for propagation. Isolated plasmids were digested by restriction enzymes (EcoRI/NcoI) and analyzed by agarose gel electrophoresis to confirm that the sequences of interest were present, and for gel purification for further downstream cloning.

pBAD-M30 cloning

pBAD-M30 vector was digested by restriction enzymes (EcoRI/NcoI) and analyzed by agarose gel electrophoresis to confirm the digestion. The digested pBAD-M30 vectors were purified for further downstream cloning. The inserts from the TA Cloning were cut, purified and cloned into pBAD-M30 vector. The pBAD-M30 clones were confirmed by dideoxy sequencing of the inserts. The pBAD-M30 constructs were heat-shock transformed into Stratagene: *E. coli* ArcticExpress™ RIL competent cells for protein expression. The transformation was performed according to the ArcticExpress™ RIL competent cells protocol.

2.2 Protein work

2.2.1 Protein expression and optimization

The protein expression was tested with different arabinose concentration (induction concentration), induction temperatures and induction lengths to optimize the yield of soluble proteins after expression.

Method

E. coli ArcticExpress™ RIL competent cells (Stratagene) were used for protein expression. The PtDDG-UNG protein expression construct were the first to be optimized, and because of limited time, the only protein expression construct to be optimized. The result of the PtDDG-UNG optimization was used with the other two protein expression constructs (PtDDG-FL and PtDDG-NEIL). Transformed cells were optimized by arabinose concentration (induction concentration): 1%, 0,2% and 0,04% arabinose, induction temperature: 10°C, 13°C, 18°C, and induction time: 20 hours, 24 hours and 26 hours.

The protein expression was performed according to the ArcticExpress™ RIL competent cells protocol. 2X YT (Appendix 1) were used instead of LB broth. OD_{600} (a spectrophotometrical proxy for cell density) was measured both before and after induction. Arabinose was used as inducer.

2.2.2 Large-scale protein expression

A large-scale protein expression protocol was made in order to get enough proteins for further downstream analysis. The results from the protein optimization were used to make the protocol.

Method

1. 3 ml 2X YT containing appropriate antibiotics (carbenicillin and gentamicin) were inoculated with a colony of *E. coli* ArcticExpress™ RIL competent cells transformed with one of the tree pBAD-M30 DDG constructs. The culture was incubated at 220 rpm and 37 °C overnight.
2. 1 ml of the culture was pipetted into an autoclaved Erlenmeyer flask containing 100 ml 2X YT (Appendix 1). The culture was incubated at 30 °C and 220 rpm for 3-4 hours, until reaching mid-log growth with an OD_{600} value of 0,5-0,8.
3. 10 ml of the culture was pipetted into a 50 ml sterile tube as a negative control. The control was centrifuged at 4000 rpm and 4 °C for 15 minutes. The supernatant was discarded and the pellet stored at -20 °C.
4. 0,2% (v/v) arabinose was added to the Erlenmeyer flask culture. The culture was incubated at 220 rpm and 18 °C for 20 hours.
5. The culture was pipetted into two 50 ml sterile tubes and centrifuged at 4000 rpm and 4 °C for 15 minutes. The supernatant was discarded and the pellet stored at -20 °C.

2.2.3 Protein separation: SDS-PAGE

The first step in SDS-PAGE is to coat the protein in sodium dodecyl sulfate (SDS). SDS unfolds and covers the proteins with a negative charge. The negative charges are proportional to the size of the proteins, because larger proteins are able to bind more negatively charged SDS molecules than smaller proteins. The proteins are separated by size by Polyacrylamide gel electrophoresis (PAGE). The proteins are then visualized by staining the gel. Coomassie and silver staining are two commonly used staining methods (Clark 2010).

Method

Gel casting

1. Bio-Rad Mini-PROTEAN® 3 Cell System was used to cast the gels. The Cell system was put together after the manufacturer's protocol. Before the gel was cast, the assembly was checked for leakage by adding distilled water to the gel cassette. The distilled water was removed by decanting and filter paper absorption.
2. The resolving gel (Appendix 1) of appropriate acrylamide percentage (7%, 10%, 15%) was prepared and poured into the gel cassette assembly, leaving enough space for the stacking gel (1-1,5 cm). Distilled water was added on top of the gel to keep it from drying. The gel was allowed to polymerize for 20-30 minutes. The distilled water was removed by decanting and filter paper absorption.
3. The stacking gel (Appendix 1) was prepared and poured on top of the resolving gel. A comb of appropriate size was inserted into the stacking gel. The gel was allowed to polymerize for 30-40 minutes.
4. Gels were inserted into the Mini Tank and the combs removed. The wells were rinsed by pouring 5X electrophoresis buffer (Appendix 1) into the Mini Tank prior to gel electrophoresis (Hem 2010).

Sample preparation and gel electrophoresis

1. Protein samples were mixed with 5X sample buffer (Appendix 1) and vortexed.
2. Samples were heated at 70 °C for 10 minutes.
3. Samples were centrifuged in a tabletop microcentrifuge at 13 000 rpm for 5 minutes at 20 °C.
4. 5 µl protein standard was loaded into the first well of the gel assembly. The samples were loaded into the polyacrylamide gel wells.
5. The gel was run at 130-150 V for 30-45 minutes (Hem 2010).

SDS-gel staining: Coomassie Brilliant Blue

Simply Blue ™ SafeStain (Invitrogen) was used to stain the SDS-gel. The staining was performed according to Invitrogen: Simply Blue ™ SafeStain microwave provided by the supplier.

2.2.4 Cell lysis: sonication

Sonication is a method that uses ultrasonic energy (sound waves) to disrupt cell membranes (Branson).

Method

A Branson 250 sonifier was used to lyse cells after protein expression was performed. Because of the difference of calculated isoelectric point (pI) of the three expressed proteins, two lysis buffers with different pH were used. Cells with PtDDG-NEIL and PtDDG-FL expression were resuspended in lysis buffer pH 8,5, and cells with PtDDG-UNG expression were resuspended in lysis buffer pH 7,0.

Branson sonifier 250 settings:

Timer: Off

Duty cycle %: 30

Output control: 3

1. A cell pellet was resuspended in cold lysis buffer (Appendix 1). 7 ml lysis buffer per gram of cell pellet was used.
2. Samples were sonicated in 6 x 30 seconds bursts with 1 minute break to prevent overheating. The sample was kept on ice during the sonification.
3. The samples were centrifuged at 25 000 g for 30 minutes at 4 °C.
4. The supernatant and the pellet were separated and the samples were stored at 4 °C (Manser 2002).

2.2.5 Cell lysis: lysozyme

Lysosyme is an enzyme that attacks the bacterial cell wall by degrading the cell wall polymer peptidoglycan (Clark 2010).

Method

1. Induced and uninduced cell pellets from protein expression cultures were resuspended in 50 µl ice-cold lysis buffer (without DTT) per 1 OD.
2. Lysozyme from chicken egg white (Sigma-Aldrich) was added to a final concentration of 1,5 mg/ml. The cell suspension was incubated on ice with rocking for 1 hour.
3. Triton X-100 was added to final concentration of 1% (v/v). The tubes were vortexed and incubated for 30 minutes with rocking.
4. DTT was added to final concentration of 5 mM.
5. Removal of insoluble debris was carried out by centrifugation at 13 000 rpm, 4 °C for 10 minutes with a tabletop centrifuge. The supernatant and pellet were separated and stored at -80 °C (Hem 2010).

2.2.6 Protein purification: Column affinity

It may often be difficult to purify and separate recombinant proteins from the host organism proteins. A solution to this problem is adding purification tags to the recombinant protein. A purification tag is a protein sequence that has high affinity binding properties for another molecule. The recombinant protein can be purified by binding the tagged protein to a solid surface, often in the form of a column matrix, and then washing away the other unbound proteins. The tagged protein can then be eluted from the column. The purity of the elution depends on the ability of other unwanted proteins to bind to the column (Reece 2009).

Method

Because of the difference of calculated isoelectric point (pI) of the three expressed proteins, two lysis (binding) and eluting buffers with different pH were used. Buffers with pH 8,5 were used with PtDDG-NEIL and PtDDG-FL protein lysate and buffers with pH 7,0 were used with PtDDG-UNG protein lysate.

His column purification

For this work, GE Healthcare HisTrap™ FF 1ml were used. A new HisTrap was used for each of the three expressed proteins.

1. The column was equilibrated with 5 x column volumes (5 ml) of MilliQ water and 5 x column volumes of lysis buffer.
2. The cell lysate was added to the column and the flow through were collected and stored at 4 °C.
3. Unbound proteins were washed by adding 5 x column volumes of lysis buffer.
4. Bound proteins were eluted by adding 5 x column volumes of Histidine eluting buffer (Appendix 1). The elution was fractioned into five eppendorf tubes and stored at 4 °C.
5. The column was washed by adding 5 x column volumes of eluting buffer, 5 x column volumes of lysis buffer, 5 x column volumes of MilliQ water. 5 x column volumes of 20% ethanol were added for storage of the column.

GST column purification

For this work GE Healthcare GSTrap™ FF 1ml was used. A new HisTrap was used for each of the three expressed proteins.

1. The column was equilibrated with 5 x column volumes (5ml) of MilliQ water and 5 x column volumes of lysis buffer.
2. The cell lysate was added to the column and the flowthrough was collected. The flowthrough was added to the column to increase the recombinant protein yield. The second flowthrough was collected and stored at 4 °C.
3. Unbound proteins were washed by adding 5 x column volumes of lysis buffer.
4. Bound proteins were eluted by adding 5 x column volumes of GST elution buffer (Appendix 1). The glutathione was dissolved in the elution buffer right before elution. The elution was fractioned into five eppendorf tubes and stored at 4 °C.
5. The column was washed by adding 5 x column volumes of eluting buffer, 5 x column volumes of lysis buffer, 5 x column volumes of MilliQ water. For storing the column 5 x column volumes of 20% ethanol was added.

2.2.7 Protein purification: Dialysis

Dialysis is the separation of larger molecules from smaller ones by selective diffusion through a semi-permeable membrane in a solute.

Method

Because of the difference of calculated isoelectric point (pI) of the three expressed proteins, two dialysis buffers with different pH were used. A buffer with pH 8,5 was used with PtDDG-NEIL and PtDDG-FL protein sample and a buffer with pH 7,0 were used with PtDDG-UNG protein sample.

1. A Spectra/Por molecular porous membrane, MWCO: 12 - 14 000, was cut in appropriate lengths according to the manufacturer's instruction (Volume/length: 0,32 ml/cm).
2. The membrane was placed in warm MilliQ water for 5 minutes in order to regenerate the membrane.
3. The membrane was clamped at one end and the sample was added into the membrane at the open end. The membrane was then clamped at the open end.
4. The membrane was incubated in 500 ml dialysis buffer (Appendix 1) for 4 hours at 4 °C by a magnet stirrer. A magnet stir bar of appropriate size was used and the stirrer was adjusted to the maximum speed were it did not pull down the membrane.
5. The buffer used for the first 4 hours was exchanged with 500 ml dialysis buffer. The membrane was incubated overnight in the new buffer, with the same conditions as described above.
6. One of the clamps was removed and the dialyzed sample was poured into an eppendorf tube. The sample was stored at 4 °C.

2.2.8 Tobacco etch virus (TEV) protease cleavage

Some proteases cleave proteins in very specific recognition sequences. The TEV protease cleaves between Gln-Gly in the protein sequence Glu-Asn-Leu-Tyr-Phe-Gln/Gly. The TEV protease can be used to remove a purification tag from a recombinant protein, by adding the TEV recognition sequence between the protein and the tag (Reece 2009).

Method

TEV protease digestion while recombinant proteins are bound to GSTrap

1. 50 µl of a 2mg/ml TEV protease stock solution were diluted in 1 ml lysis buffer.
2. The solution was added to the GSTrap, containing bound recombinant protein, before the GSTrap elution step was performed (2.2.6 Protein purification: Column affinity). The GSTrap with TEV protease solution was incubated overnight at 4 °C.
3. The GSTrap eluation step of digested proteins was performed by adding 5 x column volumes of lysis buffer. The elution was fractioned into five eppendorf tubes and stored at 4 °C.
4. Elution of undigested recombinant proteins and digested GST tags was performed by adding 5 x column volumes of eluting buffer (10 mM glutathione). The elution was fractioned into five eppendorf tubes and stored at 4 °C.
5. The TEV digestion was visualized by SDS-PAGE.

TEV protease digestion during dialysis

1. TEV protease was added during the buffer change in the dialysis method (2.2.7 Protein purification: Dialysis, step 5). At this point, the membrane was opened at one end and 75 µl of a 2mg/ml TEV protease stock solution was added to the sample.
2. The TEV digestion was verified by SDS-PAGE.

2.2.8 Purified protein quantification: Spectrophotometer

Most proteins in a solution exhibit absorbance at 280 nm (UV-light). The protein concentration can be measured by using a spectrophotometer and plotting the result against a standard curve (e.g. BSA or igG) (NanoDrop).

Method

NanoDrop® ND-1000 was used to quantify purified protein samples, using BSA as standard. The quantification was performed according to the NanoDrop® ND-1000 protocol. GST eluting buffer was used as blank. 2 µl of sample was used for the measurement (NanoDrop).

2.2.9 MALDI-TOF

Matrix-assisted laser desorption-ionization time-of flight (MALDI-TOF) mass spectrometry is a technique that generates gas-phase ions of a solid sample by using a pulsed laser. The time of flight (TOF) step measures the time an ion used from the ion source to a detector. The square root of the mass-to-charge ratio (m/z) is proportional with the TOF (Clark 2010).

Method

Purified protein samples were kept on ice and brought to the FUGE proteomics laboratory at the Department of Cancer Research and Molecular Medicine, NTNU, for MALDI-TOF analysis. The MALDI-TOF analyses were performed by principal engineer Sissel Håvåg.

2.2.10 Activity Assay: Standard UDG assay

The method is based on subjecting the enzyme to calf-thymus DNA in which radio-labeled (tritium [^{3}H]) uracil has been incorporated instead of thymine. In its double stranded form the uracils will be in the U:A base pair conformation, resembling misincorporated DNA. UDG activity results in cleavage and release of the radiolabeled uracil from the substrate and the radiolabeled free uracil bases are separated from the long DNA substrates by acid precipitation and centrifugation. The soluble free radiolabeled uracil bases are measured by liquid scintillation counting. The UDG activity correlates with the amount of counted radiolabeled uracil bases (Kavli, Sundheim et al. 2002).

Method

The assay was performed by Dr. Berit Doseth at the Department of Cancer Research and Molecular Medicine, NTNU. The assay was performed with different NaCl concentrations, $MgCl_2$ concentrations and temperatures to find the optimum conditions for PtDDG activity. In addition, the K_m and K_{cat} values of PtDDG were calculated subsequent to performing the standard assay at various substrate concentrations. The software suit EnzFitter V2 (Biosoft, UK) was used to calculate the kinetic parameters. The calculation formula used to calculate specific activity are listed in appendix 5.

1. A 20 µl mixture containing various amounts of recombinant DDG (enzyme), 5 µl [^{3}H]dUMP-labelled calf thymus DNA substrate (final 1,8 µM)and 10 µl buffer (final 0-100 mM NaCl, 20 mM Tris-HCl pH=8.0, 1 mM DTT, 1mM EDTA, 0.5 mg/ml BSA, 0-10 mM $MgCl_2$) were made. The mixture was incubated at 20-50 °C for 10 minutes.
2. The reaction was stopped by cooling on ice.
3. 50 µl salmon sperm DNA (1mg/ml in TE buffer) and 500 µl cold 5% TCA was added to the samples. The samples were gently mixed and incubated on ice for 10 minutes.
4. The samples were centrifuged in a tabletop centrifuge at 14 000 rpm for 10 minutes at 4 °C.
5. 500 µl of the supernatant were mixed with 5 ml counting fluid in a 20 ml scintillation vial. The samples were mixed by vortexing.
6. The samples were counted in a Tri-Carb 2900 TR Liquid Scintillation Analyzer (PerkinElmer) with 3 minutes counting time (Kavli, Sundheim et al. 2002).

2.2.11 Activity Assay: Oligonucleotide assay

The method is based on labeled oligonucleotide substrates to measure glycosylase activity. An active DNA glycosylase recognizes the substrate and creates an AP site. The AP site is cleaved by adding piperidine. The cleaved oligonucleotide is smaller than uncleaved oligonucleotides and can be separated by gel electrophoresis. The labeled oligonucleotides can be visualized by fluorescent imaging. The ratio between cleaved and uncleaved oligonucleotide is correlated with the glycosylase activity (Kavli, Sundheim et al. 2002).

Method

The assay was performed by Dr. Berit Doseth at the Department of Cancer Research and Molecular Medicine, NTNU. For the UNG assay 0,02 µM 19 mer U141 (5'(6FAM)-CATAAAGTGUAAAGCCTGG-3') was used as substrate in double stranded (ds) U:G, U:A and single-stranded (ss) U contexts. 0,02 µM 60 mer 5-OHU (5'-GCATGCCTGCACGG(5-oh-dU)CATGGCCAGATCCCCGGGTACCGAG-3') was used as substrate in the ss-5-OHU and ds-5-OHU NEIL assays (Wollen Steen, Doseth et al. 2012).

1. Double stranded substrate was prepared by annealing labeled oligonucleotide to unlabeled complementary strand with a G or an A base opposite to U or a G base opposite to OHU. 50% excess of the complementary unlabeled strands was used to generate the substrates (Kavli, Sundheim et al. 2002).
2. A 10 µl mixture containing 1 µl substrate, 7 µl buffer (final 10-200 mM NaCl, 20 mM Tris-HCl pH=6-9, 1 mM DTT, 1mM EDTA, 0.5 mg/ml BSA, 0-12.5 mM $MgCl_2$) and various amounts of recombinant DDG (enzyme) was made. The mixture was incubated at 37 °C for 20 minutes.
3. 50 µl 10% piperidine was added and the samples were heated at 90 °C for 20 minutes.
4. The samples were vacuum dried (speedvac) at 60 °C for 1 hour.
5. The samples were separated by SDS-PAGE (12%) and visualized by laser scanning (Typhoon Trio Imager). The samples were analyzed by using imageQuant TL software (GE Healthcare) (Kavli, Sundheim et al. 2002).

3. Results and discussion

3.1 Cloning of protein expressing constructs

An NcoI restriction site in the *PtDDG* UNG domain was removed from the *PtDDG* template (*PtDDG* in pENTR vector) by QuickChange mutagenesis before the cloning steps were started. The mutagenesis was verified by restriction enzyme digestion and agarose gel electrophoresis. Lane 1 and 2 of figure 3.1 shows that the vector has not been digested by NcoI and suggest that the mutagenesis has been successful. The lane 3 sample did not look as promising as the previous two samples and might be digested by NcoI at some degree. Lane 4 shows that the unmutated control has been digested by NcoI and forms a band that correlates with the expected size (1058 bp) between two NcoI restriction sites in the vector and *PtDDG*. The samples of lane 1 and 2 were further verified by DNA sequencing.

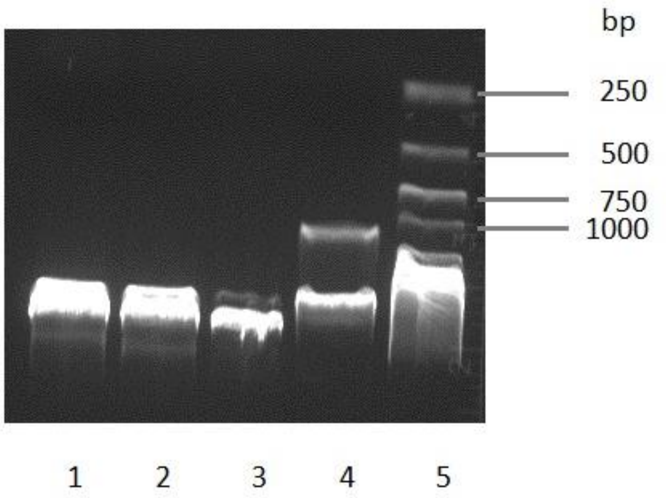


Figure 3.1: Agarose gel electrophoresis of NcoI digested QuickChange mutated PtDDG. Lane 1,2,3: mutated *PtDDG*, lane 4: unmutated *PtDDG*, lane 5: GeneRuler™ 1kb DNA ladder (Fermentas).

A vector containing mutated *PtDDG* was used for PCR to produce the sequences *PtDDG-FL*, *PtDDG-NEIL* and *PtDDG-UNG* with primers introducing a 5' NcoI restriction sites and a 3' EcoRI site (Appendix 3). The *PtDDG-FL*, *PtDDG-NEIL* and *PtDDG-UNG* PCR products were subcloned into a TA cloning vector to obtain sufficient amounts of DNA for further downstream cloning, and then cloned into the expression vector pBADM-30 (Figure 3.2) using the restriction enzymes NcoI and EcoRI.

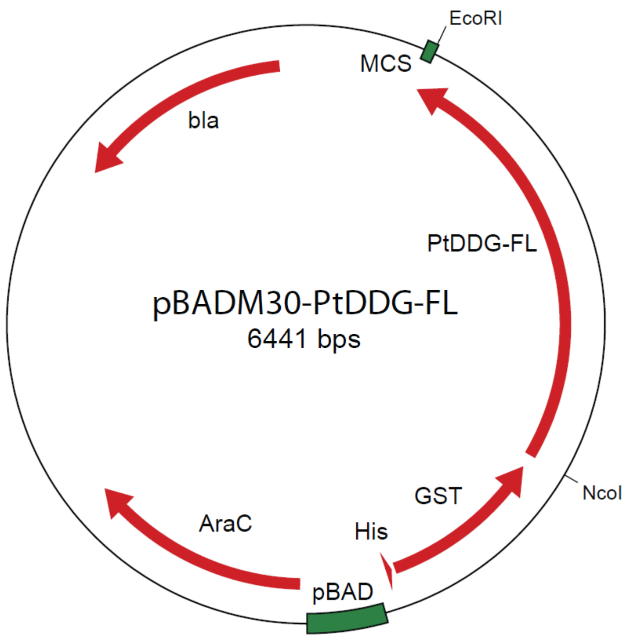


Figure 3.2: pBADM30 PtDDG-FL plasmid map. Schematic overview of the key features of the pBADM30 expression vector cloned with PtDDG-FL.

The pBADM-30 constructs were digested with the same restriction enzymes and analyzed by agarose gel electrophoresis to verify that the *PtDDG* sequences had been cloned into the vector. Figure 3.3 shows that the length of the digested expression constructs were of expected size: *PtDDG-NEIL* length: 1020 bp, *PtDDG-UNG* length: 651 bp, *PtDDG-FL* length: 1794 bp and pBAD-M30 vector length: 4755 bp. The pBAD-M30 expression constructs were further verified by DNA sequencing.

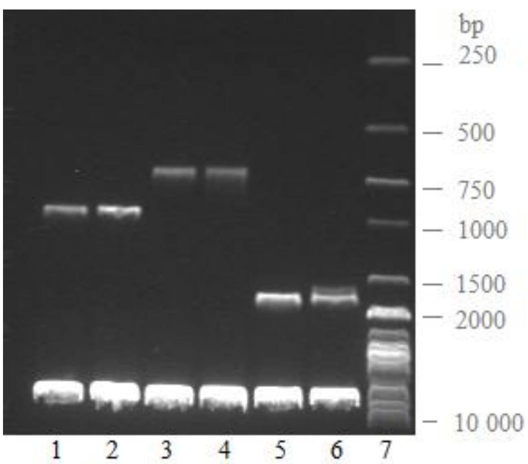


Figure 3.3: Verification of pBAD-M30 recombinant protein constructs. Agarose gel electrophoresis of pBAD-M30 vector cloned with PtDDG-NEIL domain, PtDDG-UNG domain and PtDDG-FL, which has been digested with EcoRI/NcoI restriction enzymes. Lane 1,2: PtDDG-NEIL, 3,4: PtDDG-UNG, 5,6: PtDDG-FL, 7: GeneRuler™ 1kb DNA ladder (Fermentas).

3.2 Expression of soluble PtDDG-FL, PtDDG-UNG and PtDDG-NEIL

Recombinant PtDDG-FL, PtDDG-UNG and PtDDG-NEIL were transformed into and expressed in *E. coli* ArcticExpress™ cells by induction with arabinose (0,2%). Two different cell lysis methods were attempted in order to obtain soluble recombinant proteins. Cell lysis with lysozyme did not result in any visible soluble fraction (lysate) of expressed recombinant proteins. Several attempts were made after changing the pH of the lysis buffer in correlation with the predicted pI of the proteins (PtDDG-FL pI: 7,41, PtDDG-UNG pI: 8,03, PtDDG-NEIL pI: 6,74**).** Figure 3.4 shows that the recombinant proteins have been expressed, but are only seen in the pellet fractions after lysozyme lysis. The absence of recombinant proteins in the soluble fraction might be because of formation of inclusion bodies (protein aggregates).

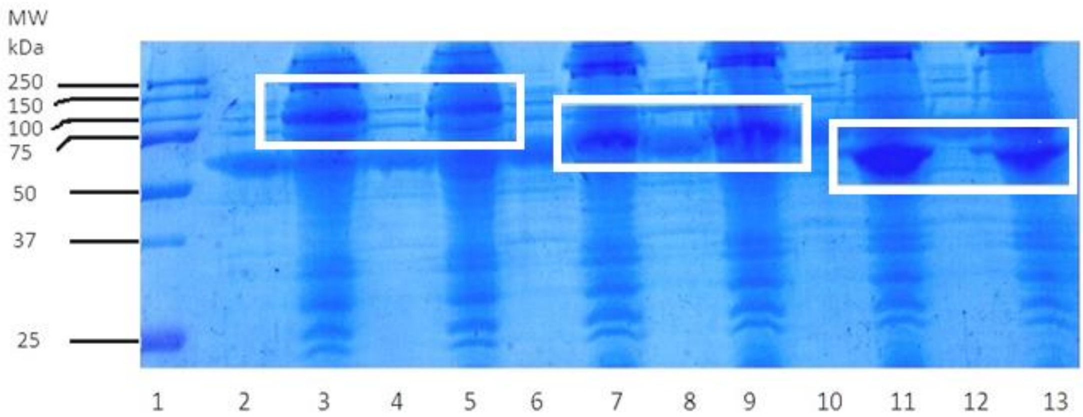


Figure 3.4: Production of soluble PtDDG by lysozyme treatment. SDS-PAGE of induced cells lysed by lysozyme. Lane 1: Precision plus protein ladder (5 µl, BioRad, all blue), lane 2,4: PtDDG-FL lysate, lane 3,5: PtDDG-FL pellet, lane 6,8: PtDDG-NEIL lysate, lane 7,9: PtDDG-NEIL pellet, lane 10,12: PtDDG-UNG lysate, lane 11,13: PtDDG-UNG pellet. The white boxes show the expressed recombinant proteins.

Sonication of cells was performed in order to get soluble recombinant proteins. Figure 3.5 shows the SDS-PAGE results of sonicated PtDDG-UNG expressing cells.

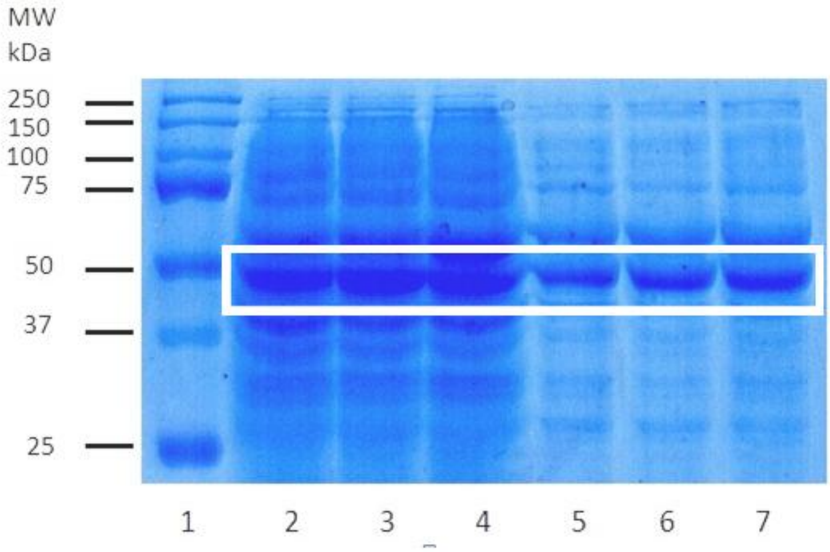


Figure 3.5: Production of soluble PtDDG by lysozyme treatment. SDS-PAGE of induced cells lysed by sonication. Lane 1: Precision plus protein ladder (5 µl, BioRad, all blue), lane 2,3,4: PtDDG-UNG lysate, lane 5,6,7: PtDDG-UNG pellet. The white square shows the recombinant PtDDG-UNG.

The figure shows that most of the cells proteins are in the soluble fraction (lane 2, 3 and 4), including the outlined recombinant protein PtDDG-UNG (predicted molecular weight of 54 kDa). Similar results were seen in the other two protein expressing cells (PtDDG-FL and PtDDG-NEIL). This result is very different from the lysozyme-treated cells, where almost all the proteins were seen in the pellet (non-soluble fraction).

Comparison of the results from sonication and lysozyme treatment strongly suggests that sonication was the most suited method for getting soluble recombinant proteins after expression.

3.3 Optimization of recombinant PtDDG-UNG expression

Once we had selected a proper protocol for cell lysis, we started to make an optimization protocol for PtDDG-UNG. The PtDDG-UNG protein expression construct was the first to be optimized, and because of limited time, the only protein expression constructs to be optimized. Transformed *E. coli* ArcticExpress ™ cells were optimized with regard to arabinose concentration (0,04%, 0,2 % and 1% arabinose induction concentration), induction temperature (10 °C, 13 °C and 18 °C) and induction time (20, 24 and 26 hours). Figure 3.6

shows the results after 20 hours incubation time, with different arabinose concentrations (0,04%, 0,2% and 1%) and temperatures (10 °C, 13 °C and 18 °C). The samples are from the soluble fraction (lysate) after sonication.

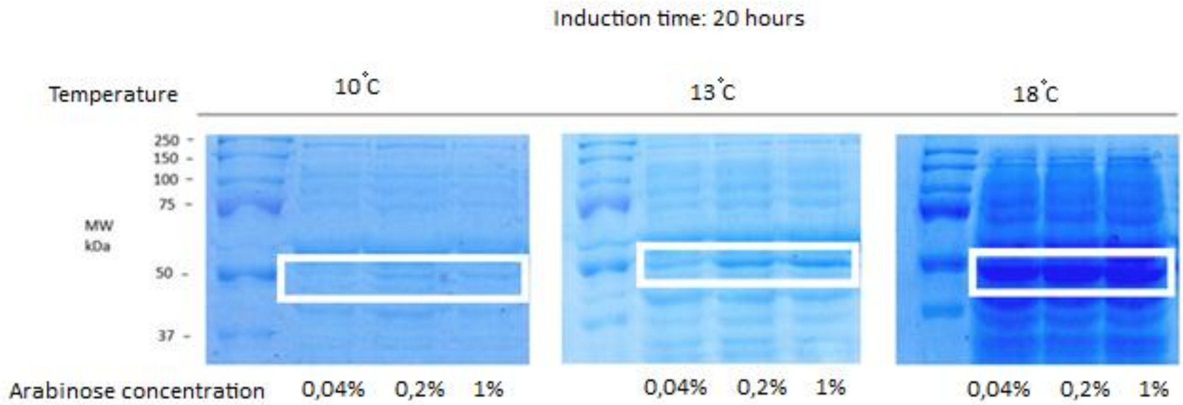


Figure 3.6: Expression optimization of PtDDG-UNG at 20 hours incubation, with different arabinose concentrations and temperatures. All samples are from soluble fraction (lysate) after sonication, and the same amount of sample volume (25 µl) is used at each lane. Precision plus protein ladder was applied in the left lane of each of the SDS-PAGE pictures (5 µl, BioRad, all blue). The white squares indicate the expected size of PtDDG-UNG.

The figure clearly shows a difference in the amount of recombinant proteins in the soluble fractions. At 10 °C and 13 °C, the expressed recombinant protein can barely be seen on the gel. At 18 °C the amount of recombinant protein is much more noticeable. The differences of expressed recombinant proteins because of arabinose concentrations (induction concentration) are not very obvious. It seems that slightly more recombinant proteins are expressed at arabinose concentration of 0,2 % compared to 0,04 %. The amount of proteins expressed at arabinose concentrations of 0,2 % and 1 % appears to be identical. Different induction times (20, 24 and 26 hours) did not result in any noticeable differences in recombinant protein yield. To summarize the results from the PtDDG-UNG optimization, it seems that the optimum parameters to gain most soluble recombinant protein are as follows: arabinose concentration: 0,2%, induction temperature: 18 °C and induction time: 20 hours. The results from the PtDDG-UNG optimization were also used with the other two protein expression constructs (PtDDG-FL and PtDDG-NEIL).

3.4 Purification of recombinant PtDDG-FL, PtDDG-UNG and PtDDG-NEIL

After protein expression of recombinant PtDDG-FL, PtDDG-NEIL and PtDDG-UNG in *E. coli* ArcticExpress™ cells and cell lysis was performed, the proteins were purified by histidine tag and GST tag affinity chromatography. Figure 3.7 shows PtDDG-UNG lysate purified by histidine tag affinity chromatography.

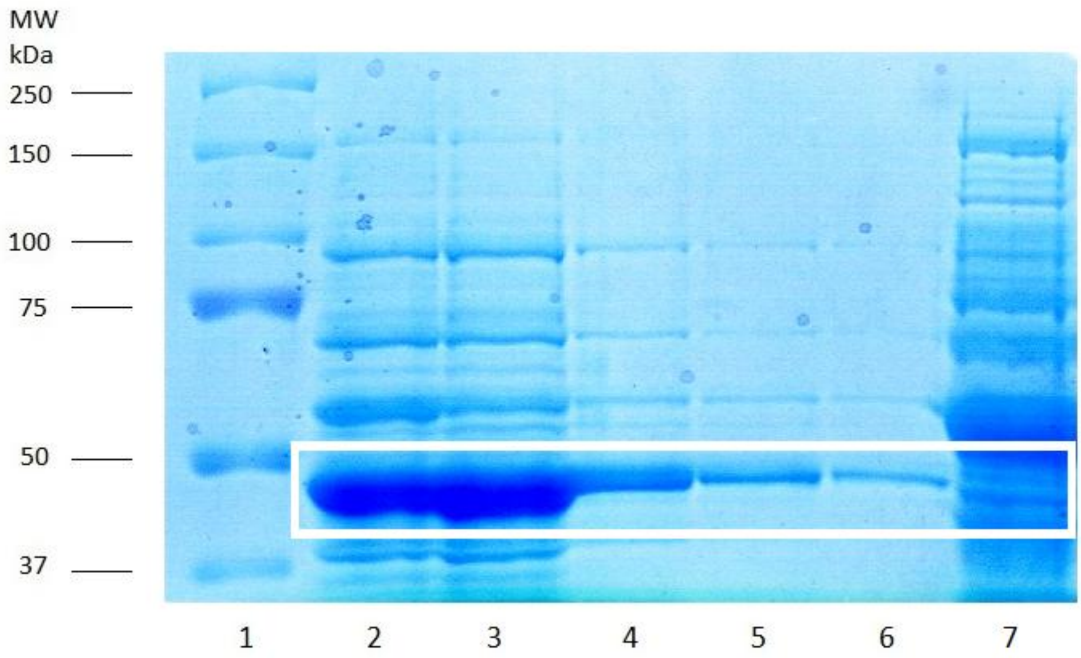


Figure 3.7: Purification of PtDDG-UNG by His tag affinity chromatography. SDS-PAGE of five elutions of His tag column purified recombinant PtDDG-UNG**.** Lane 1: Precision plus protein ladder (5 µl, BioRad, all blue), lane 2-6: first to fifth elution fraction of His-tag column purified recombinant PtDDG-UNG, lane 7: flow trough. The white square indicates the expected size of the recombinant PtDDG-UNG.

Similiar results were also seen with the two other protein expression construct (PtDDG-FL and PtDDG-NEIL), but because PtDDG-UNG was the only expression construct to be optimized, the amount of produced protein was somewhat larger. Comparison of the purified fractions (lanes 2-6) in figure 3.7 with the flowthrough fraction (lane 7) shows that the purity of the sample was increased after the histidine tag purification step. Nevertheless, there were still some unwanted proteins in the sample. It is also notable to see that most of the expressed recombinant PtDDG-UNG have been eluted in the purified fraction and are only seen as a faint band in the flowthrough.

The samples were further purified by dialysis and by a GST-tag purification step. Figure 3.8 shows PtDDG-FL, PtDDG-NEIL and PtDDG-UNG purified by dialysis and GST-tag affinity chromatography.

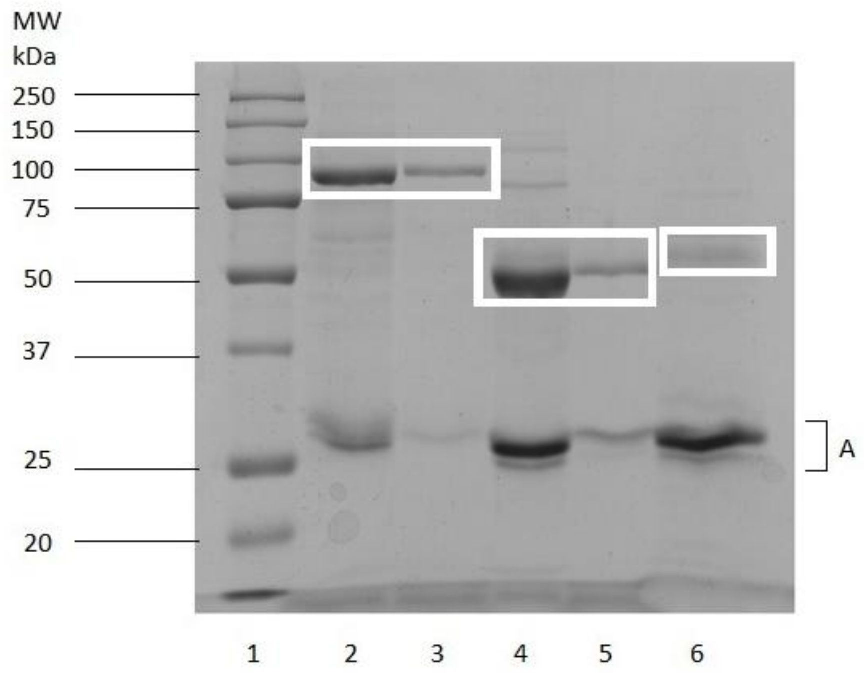


Figure 3.8: Second step purification of PtDDG-FL, PtDDG-UNG and PtDDG-NEIL by GST-tag affinity chromatography. SDS-PAGE separation of recombinant PtDDG-FL, PtDDG-UNG and PtDDG-NEIL purified by histidine-tag affinity chromatography, dialysis and GST-tag affinity chromatography. Lane 1: Precision plus protein standard (5 µl, BioRad, all blue), lane 2: PtDDG-FL first elution, lane 3: PtDDG-FL second elution, lane 4: PtDDG-UNG first elution, lane 5: PtDDG-UNG second elution, lane 6: PtDDG-NEIL (concentrated sample) elution. A: Produced empty GST/His-tag. The white squares show the purified recombinant proteins.

Comparison of figure 3.7 with figure 3.8 shows that most of the contaminating bacterial proteins have been removed from the samples, but some faint bands can still be seen in the first elution lanes and the concentrated PtDDG-NEIL lane. The figure also shows that empty GST/histidine-tags (A) have been produced and purified in all the samples. Because of unexpected low recombinant protein gain in the PtDDG-NEIL sample, it was concentrated by a protein concentrate spin column (Amicon® Ultra-4 centrifugal filter unit, 30 000 MWCO, Millipore), but the amount of recombinant PtDDG-NEIL was still lower compared to recombinant PtDDG-UNG and PtDDG-FL. Because of the concentrating step, more empty GST/histidine-tags are seen in the PtDDG-NEIL lane.

TEV protease was used in an attempt to remove the rather large GST/histidine-tag (≈26 kDa) from the recombinant proteins. For some reason, the TEV protease did not lead to any TEV site digestion on the recombinant proteins. This may be because the recombinant proteins have blocked the TEV protease site and hindered the TEV protease to get access to the TEV site. TEV protease from the same batch was used in another experiment and under similar conditions the same day, and was proven functional. The TEV protease was used both while the recombinant proteins were bound in the GST column and later during the dialysis purification step. SDS-PAGE of digested proteins did not show any visible signs of TEV site digestion.

3.5 MALDI-TOF identification of recombinant PtDDG-FL, PtDDG-UNG and PtDDG-NEIL

The purified samples were analyzed by MALDI-TOF to confirm the presence of expressed recombinant proteins. The results (Appendix 4) verified the presence of PtDDG-UNG (*Phaeodactylum tricornutum* CCAP 1055/1) in both the PtDDG-UNG sample and in the PtDDG-FL sample. The PtDDG-NEIL sequence is not correctly added to any database and was verified by MSMS sequence analysis and manual peptide matching by using our own annotated version of the PtDDG-NEIL amino acid sequence. PtDDG-NEIL was verified in both the PtDDG-NEIL sample and in the PtDDG-FL sample.

3.7 Oligonucleotide assays of recombinant PtDDG-FL

Purified recombinant PtDDG-FL was used in oligonucleotide assays to determine activity against ss and ds UDG substrate U141 and ss and ds NEIL substrate 5-hydroxyuracil (5-OHU). Figure 3.9 shows that the recombinant PtDDG-FL has activity against both dsDNA and ssDNA-U141 substrate. The -piperidine results, which are used as a negative control, confirm that the substrates are not cleaved under conditions where piperidine is not added to the reaction.

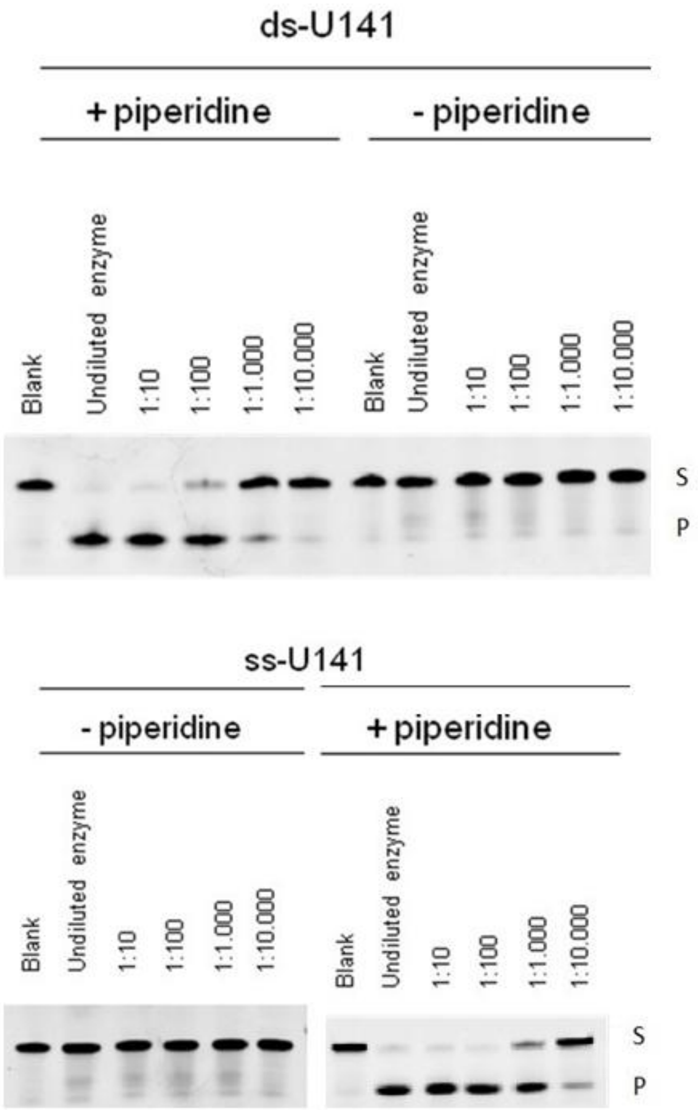


Figure 3.9: Enzyme activity of PtDDG-FL against double stranded (ds) and single stranded (ss) U141 substrate (UDG substrate). SDS-PAGE of oligonucleotide assays**.** Amount of recombinant PtDDG-FL used per ds assay reaction: 10 ng. Amount of recombinant PtDDG-FL used per ss assay reaction: 1 ng. - piperidine is used as a control. S = substrate. P = product (cleaved substrate).

Figure 3.9 also indicates that recombinant PtDDG-FL has higher activity against ssDNA substrate than dsDNA substrate. This correlates with the result seen from recombinant human UNG activity assays (Slupphaug, Eftedal et al. 1995). Cytosine deamination occurs at an 300-fold faster rate in ssDNA compared to dsDNA (Lindahl 1993). The preference for ssDNA can be of critical biological importance for maintaining the DNA (Hagen, Kavli et al. 2008).

Figure 3.10 shows a recombinant PtDDG-FL NEIL activity assays against both ds and ss-5-OHU substrate.

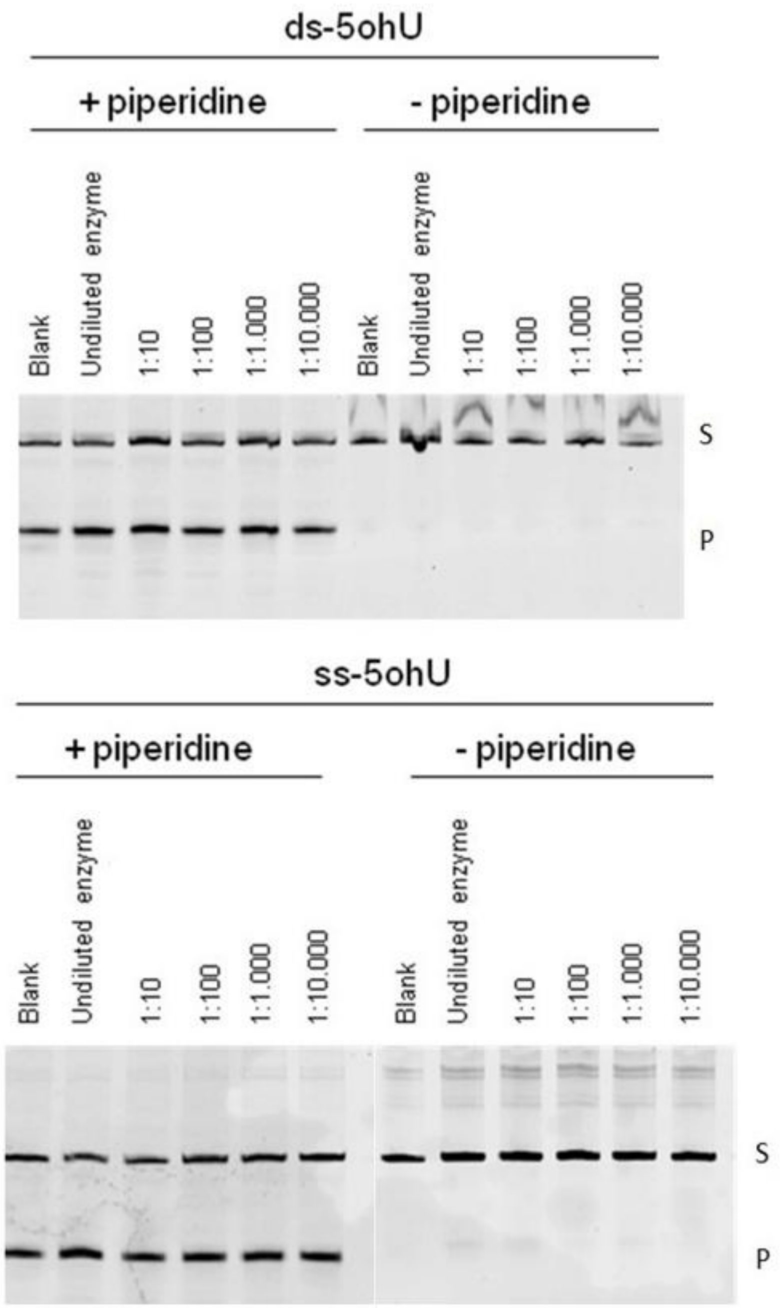


Figure 3.10: Enzyme activity of PtDDG-FL against ds and ss 5-OHU substrate (NEIL substrate). SDS-PAGE of oligonucleotide assays. Amount of recombinant PtDDG-FL used per ds and ss assay reaction: minimum 2 ng. - piperidine is used as a control. S = substrate. P = product (cleaved substrate).

All samples with added piperidine were similar to the blank, indicating the substrate was cleaved even when the enzyme was not present. The result may be caused by defects in the 5-OHU substrate. The reason for this might be because some of the 5-OHU oligonucleotides have lost their labeled base and have a pre-created AP site. The same results will then be seen in a blank sample as a sample containing recombinant PtDDG-FL. Figure 3.10 also shows a vague band in the undiluted and in the 1:10 dilution sample against ss-5-OHU without the addition of piperidine, indicating that PtDDG-FL have a weak NEIL activity against the substrate. Another known NEIL substrate, thymine glycol (Tg), has been ordered to be able to assess NEIL activity further in PtDDG-FL. Because of limited time, assays with the new substrate will not be performed in time to be included in this thesis.

An oligonucleotide assay with different substrates (ssU, U:A and U:G U141) was performed in order to analyze the substrate preference of recombinant PtDDG-FL. Figure 3.11A) shows the SDS-PAGE results of the assay and figure 3.11B) shows a comparison of the results gained from the assay.

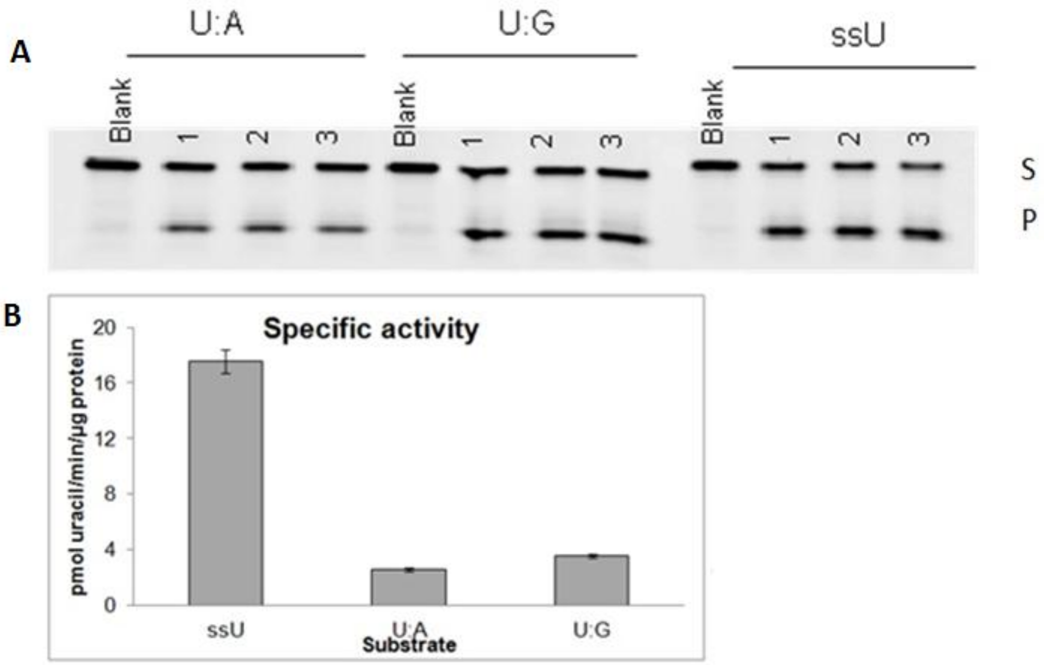
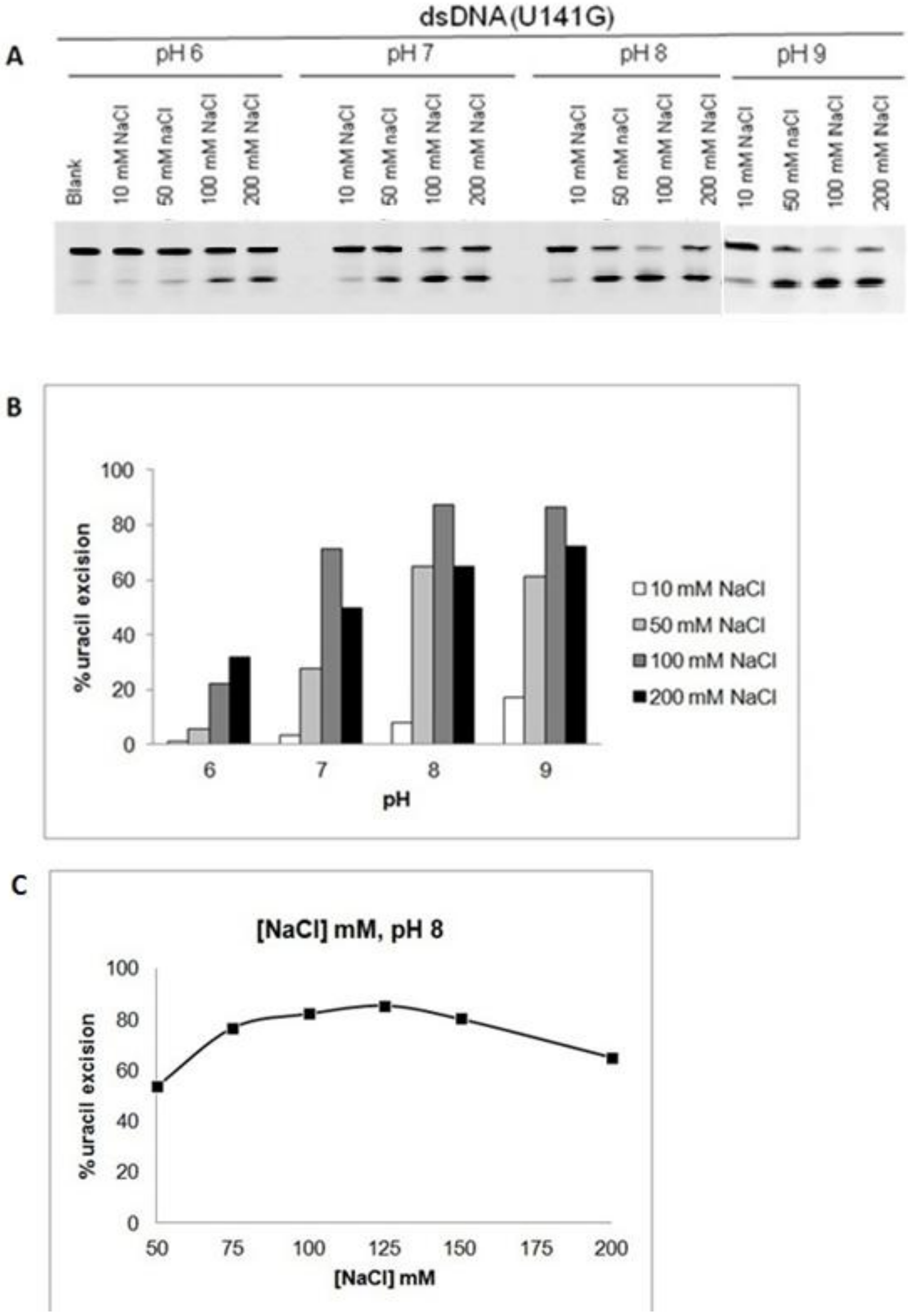


Figure 3.11: A) Enzyme activity of PtDDG-FL against U:A, U:G and ssU U141 substrate (UDG substrate). SDS-PAGE of oligonucleotide assays**.** Amount of recombinant PtDDG-FL used per U:A and U:G assay reaction: 1,33 ng. Amount of recombinant PtDDG-FL used per ssU reaction: 0,33 ng. - piperidine is used as a control. S = substrate. P = product (cleaved substrate). The assay was done with three parallels (1-3). B) Comparison of results from PtDDG-FL ssU, U:A and U:G substrate preference oligonucleotide assays. PtDDG-FL specific activity against the three substrates. Activity is given in pmol uracil released per minute per µg protein.

The results suggest that recombinant PtDDG-FL has the substrate preferences ssU>U:G and a slight preference for U:G>U:A. This is the same preference as previously shown for human UNG2 (Slupphaug, Eftedal et al. 1995). As stated earlier, the preference for ssDNA can be crucial, as deamination of ssDNA in e.g. replication forks can lead to mutations if left unrepaired (Hagen, Kavli et al. 2008), and because cytosine deamination occurs at a much higher rate ssDNA compared to dsDNA (Lindahl 1993). PtDDG-FL specific activity was found to be 17,6 pmol uracil per minute per µg protein against ssU, 2,6 uracil per minute per µg protein against U:A and 3,5 pmol uracil per minute per µg protein against U:G. Notably, if excised from ssDNA at replication forks, however, the resulting AP-site would have to await further processing until the dsDNA conformation is restored by fork regression or recombination to avoid generation of double-strand breaks and to ensure the presence of a template during repair incorporation.

An oligonucleotide assay with different pH and NaCl concentrations was performed in order to identify recombinant PtDDG-FL pH and NaCl optimum. Figure 3.13A shows the SDS-PAGE results of the assay against the substrate U141G (UDG substrate). Figure 3.13B shows a result comparison from the same assay. Figure 3.13C shows the result from an oligonucleotide assay to further identify recombinant PtDDG-FL NaCl optimum. The figure indicates that recombinant PtDDG-FL optimum is at a NaCl concentration of 125 mM and pH 8. Recombinant PtDDG-FL has a notably higher NaCl optimum compared with human UNG2, which is between 40-60 mM (Doseth, Ekre et al. 2012). The reason for a higher NaCl optimum for PtDDG-FL might be explained by the normal habitat of the organisms. *P. tricornutum* is normally found in sea water which has higher NaCl concentration than human cells. The PtDDG pH optimum of 8 is almost the same as described for human UNG2 and human SMUG1, which have a pH optimum of 7-7,5 (Kavli, Sundheim et al. 2002). The somewhat higher optimum of PtDDG-FL might be explained by the higher pH of seawater compared to the cytosolic pH of human cells.



Figur 3.13: A) Identification of pH and NaCl optima for PtDDG. SDS-PAGE of oligonucleotide assays of recombinant PtDDG-FL against dsDNA (U141G) substrates (UDG substrate) at different pH and NaCl concentrations. Amount of recombinant PtDDG-FL used per ds and ss assay reaction: minimum 2 ng. -piperidine is used as a control. S = substrate. P = product (cleaved substrate).B) Comparison of results from PtDDG-FL pH and NaCl optimum oligonucleotide assays. The figure shows percentage uracil excision under different NaCl and pH concentrations. C) Oligonucleotide assays of recombinant PtDDG-FL acitivity against U141 U:G substrate (UDG substrate) at pH 8 and with different concentrations of NaCl. Amount of recombinant PtDDG-FL used per reaction: 10 ng.

3.8 Standard UDG assay of recombinant PtDDG-FL

A standard UDG assay with different $MgCl_2$ concentrations was performed to find recombinant PtDDG-FL $MgCl_2$ optimum (Figure 3.14).

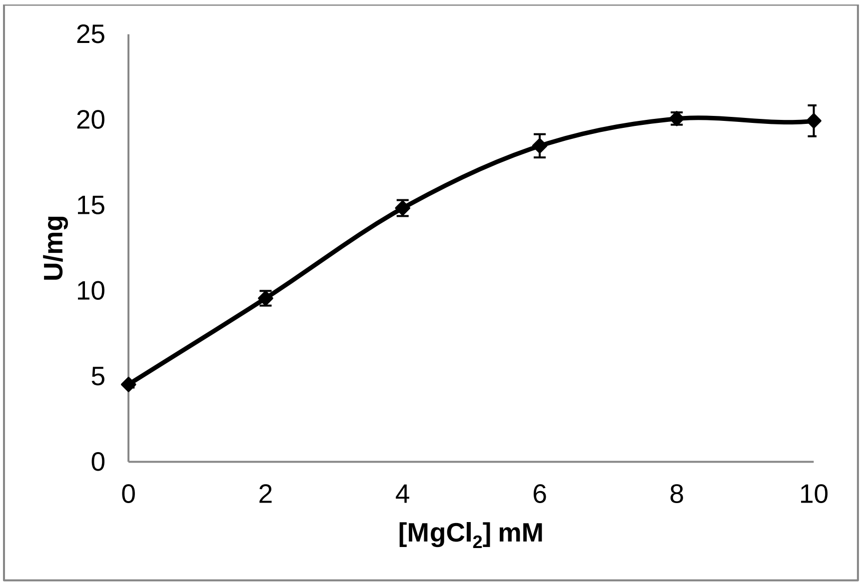


Figure 3.14: Identification of $MgCl_2$ optimum for PtDDG. Standard UDG assay of recombinant PtDDG-FL acitivity against [^{3}H]dUMP-labelled calf thymus DNA (long U:A(UDG substrate)) in different concentrations of $MgCl_2$. Amount of recombinant PtDDG-FL used per reaction: 20 ng.

The results seen from figure 3.14 indicate that the enzyme activity of recombinant PtDDG-FL is stimulated by $MgCl_2$ and that the activity is at a peak at 8 mM $MgCl_2$.The result are in between the reported optimums for human UNG2, which peaks at 6 and 10 mM $MgCl_2$ for dsU and ssU, respectively (Kavli, Sundheim et al. 2002).

Another standard UDG assay with recombinant PtDDG-FL was performed in order to find the optimum temperature of PtDDG-FL (Figure 3.15).

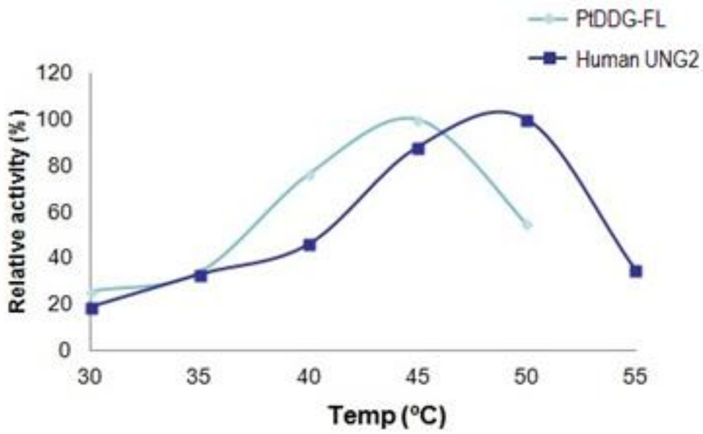


Figure 3.15: Identification of temperature optimum for PtDDG. Standard UDG assay of recombinant PtDDG-FL to determine temperature optimum. Substrate used: [^{3}H]dUMP- labeled calf thymus DNA (long U:A). Amount of recombinant PtDDG-FL used per reaction: 6,7 ng. Human UNG2 values are added for comparison (Torseth, Doseth et al. 2012).

Figure 3.15 shows that recombinant PtDDG-FL has a temperature optimum at 45 °C, which is 5 °C lower than for human UNG2 (Torseth, Doseth et al. 2012). It is notable to see that the curve shown in figure 3.15 for recombinant PtDDG-FL and human UNG2 looks very similar, only with a 5 °C displacement. To determine the thermal stability of recombinant PtDDG-FL at the enzyme temperature optimum, another standard UDG assay was performed (Figure 3.16).

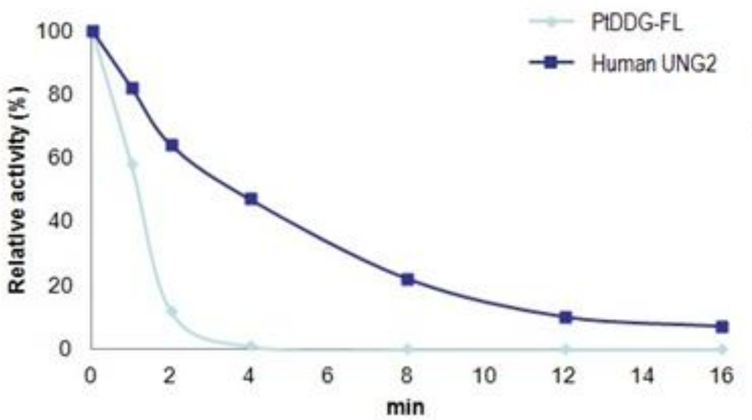
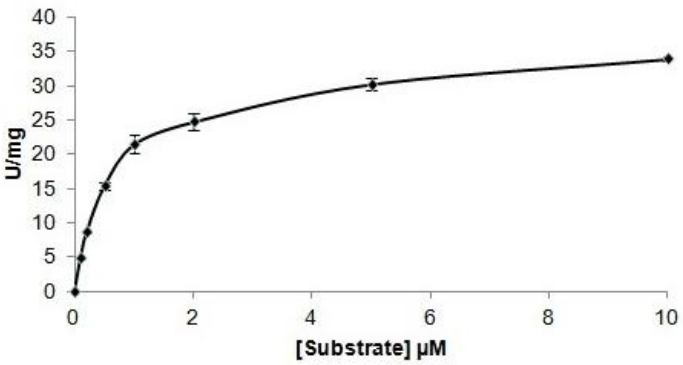


Figure 3.16: Thermal stability of PtDDG-FL at its temperature optimum. A standard UDG assay of recombinant PtDDG-FL and human UNG2 was performed at 45°C (PtDDFG-FL) and 50°C (HsUNG2). Substrate used: [^{3}H]dUMP- labeled calf thymus DNA (long U:A). Amount of recombinant PtDDG-FL used per reaction: 20 ng. Human UNG2 values are added for comparison (Torseth, Doseth et al. 2012).

The results seen from figure 3.16 indicate that recombinant PtDDG-FL activity drops fast at its temperature optimum. The relative activity is halved after approximately 1 minute and no activity is measured after 4 minutes. This differ greatly from the human UNG2 thermal stability, which halves the relative activity after approximately 4 minutes and show activity for more than 16 minutes under the same conditions (temperature optimum at 50 °C) (Torseth, Doseth et al. 2012).

To be able to calculate K_m and K_{cat} of recombinant PtDDG-FL, a standard UDG assay against [^{3}H]dUMP-labeled calf thymus DNA (long U:A) substrate was performed (Figure 3.17). The results from the assay show how active PtDDG-FL is at different substrate concentrations.



[Substrat] μM	U/mg	STDEV
0	0	0
0,1	4,8	0,2
0,2	8,5	0,1
0,5	15,3	0,6
1	21,4	1,3
2	24,7	1,2
5	30,2	0,9
10	33,9	0,4

Figure 3.17: Standard UDG assay of recombinant PtDDG-FL to determine Km/Kcat. Substrate used: [^{3}H]dUMP-labeled calf thymus DNA (long U:A). Amount of recombinant PtDDG-FL used per reaction: 20 ng.

The specific activities measured (figure 3.17) at the different substrate concentrations were used to calculate the K_m and K_{cat} values of recombinant PtDDG-FL. The K_m and K_{cat} values of PtDDG-FL against double-stranded uracil-containing DNA were 0.64 μM and 3.3 min^{-1}, respectively. Interestingly, the corresponding values for the nuclear human UNG2 enzyme are 3.0 μM and 187 min^{-1}, respectively. Thus PtDDG has nearly 5-fold higher affinity for the substrate than human UNG2, while the enzymatic turnover is about 57-fold lower.

It is tempting to speculate that these significant differences in kinetic behavior might reflect the highly specialized function of the human enzyme to detect and rapidly excise misincorporated uracils at the replication fork. Here UNG2 is apparently tethered to the DNA polymerase processivity factor PCNA by way of its N-terminal PCNA-interacting motif (PIP-motif) (Otterlei, Warbrick et al. 1999). UNG2 would thus be ideally situated to "find" misincorporated uracils as they exit the replicating polymerase even at an apparently low K_m of 3.0 µM. In contrast, PtDDG does not contain a PCNA-interacting motif, and conceivably would rely more on random collisions with the substrate and a lover K_m for catalysis to occur. The co-operation of UNG2 with the replicative machinery could also explain its markedly higher turnover rate than PtDDG and that facilitates excision of misincorporated uracil without slowing the rapidly moving replication fork.

Further analysis of the PtDDG enzyme will reveal to what extent the enzyme co-operates with the Diatome replication fork, or whether the enzyme has evolved functionalities distinct from UNG2.

3.9 Intracellular localization of PtDDG-FL and PtUNG1

An earlier master student (Cecile Anne Dahl Hem) cloned *PtDDG-FL* and *PtUNG1* into the pDEST-N-EYFP vector (Siaut, Heijde et al. 2007). This vector facilitates expression of N-terminal EYFP protein fusions in *P. tricornutum* for intracellular localization of proteins. The YFP fusion vectors were used for biolistic transformation of *P. tricornutum.* Two brownish colonies were seen on the PtUNG1 selective (zeocin) f/2 agar plates, suggesting that the transformation was successful. Promising colonies were grown in growth media both with and without selection (zeocin). When analyzed by confocal microscope fluorescence imaging, only autofluorescence from chloroplasts was observed. Further investigation of why the transformation failed lead to the discovery that a co-transformation plasmid had not been included in the transformation protocol. The missing co-transformation plasmid contains the *sh ble* gene which confer resistance to zeocin. The co-transformation plasmid is crucial for successful transformation and should have been transformed with the EYFP-fusion protein in a 1:1 ratio. Because the biolistic transformation procedure is a time consuming process (4-7 weeks) and because of own time restrictions the co-transformation was not performed.

4. Conclusion

The gene encoding "Dual DNA glycosylase" from the diatom *P. tricornutum* was cloned into the expression vector pBADM-30 as *PtDDG-FL*, *PtDDG-UNG* and *PtDDG-NEIL*. The clones were verified by sequencing and transformed into *E. coli* ArticExpress™ RIL cells. Cells containing expressed recombinant proteins were sonicated in order to get soluble recombinant proteins. Recombinant protein expression was optimized by induction time, temperature and inducer concentration cells. Soluble recombinant proteins were purified by both histidine-tag and GST-tag affinity chromatography and verified by MALDI-TOF analysis.

Recombinant PtDDG-FL was used in UDG activity assay analyses to determine the enzyme UDG activity optima for different parameters. The pH optimum was found to be at pH 8, NaCl optimum at 125 mM, $MgCl_2$ optimum at 8 mM $MgCl_2$ and temperature optimum at 45 °C. PtDDG-FL substrate preference was found to be ssU>U:G>U:A; the specific activities were 17,6 pmol uracil per minute per µg protein against ssU, 2,6 uracil per minute per µg protein against U:A and 3,5 pmol uracil per minute per µg protein against U:G. The thermal stability of PtDDG-FL at its optimum temperature was found to be low in comparison to human UNG2 results. The K_m and K_{cat} values of PtDDG-FL against double-stranded uracil-containing DNA were measured and calculated to be 64 µM and 3.3 min^{-1}, respectively. The corresponding K_m and K_{cat} values for the nuclear human UNG2 enzyme are 3.0 µM and 187 min^{-1}, respectively.

Biolistic transformation of *P. tricornutum* was performed in order to visualize the intracellular localization of PtUNG1 and PtDDG-FL. The transformation failed because of a missing co-transformation vector.

5. Recommendations for further work

The results from the PtDDG-UNG optimization were used to express recombinant PtDDG-FL and PtDDG-NEIL. In order to increase the amount of soluble recombinant PtDDG-FL and PtDDG-NEIL in further work, it could be beneficial to optimize their protein expression individually. Especially for recombinant PtDDG-NEIL expression, that seems to be poorly expressed when using the optimized PtDDG-UNG settings.

Removal of the GST/histidine-tag from the recombinant proteins is also recommended, since it's rather large size might have an effect on the proteins activity.

It would have been interesting to perform activity assays with the single domain recombinant proteins produced, PtDDG-UNG and PtDDG-NEIL, to be able to compare them with the results of PtDDG-FL and other UDG and NEIL enzymes. Other NEIL substrates should also be tested, to be able to find PtDDG-NEIL substrate preference and compare the results with other known NEIL enzyme preferences. A more extensive activity study and crystallographic studies of PtDDG-FL are also recommended for further work.

In order to visualize the intracellular localization of PtUNG1 and PtDDG-FL, the co-transformation plasmid containing the *sh ble* gene (conferring resistance to zeocin) should be used to successfully transform *P. tricornutum* with YFP fusion vector.

References

Alberts, J., Lewis,Raff, Roberts, Walter (2008). Molecular Biology of the Cell, Garland Science.

Armbrust, E. V. (2009). "The life of diatoms in the world's oceans." Nature 459(7244): 185-192.

Bio-Rad Instruction Manual: Biolistic® PDS-1000/He Particle Delivery System

Bowler, C., A. E. Allen, et al. (2008). "The Phaeodactylum genome reveals the evolutionary history of diatom genomes." Nature 456(7219): 239-244.

Bradbury, J. (2004). "Nature's nanotechnologists: unveiling the secrets of diatoms." PLoS Biology 2(10): e306.

Branson "Instruction Manual: 250 - 450 Sonifier Analog Cell Disruptor User’s Manual."

Chen, C. Y., D. W. Mosbaugh, et al. (2004). "Mutational analysis of arginine 276 in the leucine-loop of human uracil-DNA glycosylase." The Journal of Biological Chemistry 279(46): 48177-48188.

Clark, D. P. (2010). Molecular Biology, AP Cell.

Darwanto, A., J. A. Theruvathu, et al. (2009). "Mechanisms of base selection by human single-stranded selective monofunctional uracil-DNA glycosylase." The Journal of Biological Chemistry 284(23): 15835-15846.

Doseth, B., C. Ekre, et al. (2012). "Strikingly different properties of uracil-DNA glycosylases UNG2 and SMUG1 may explain divergent roles in processing of genomic uracil." DNA Repair.

Doublie, S., V. Bandaru, et al. (2004). "The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity." Proceedings of the National Academy of Sciences of the United States of America 101(28): 10284-10289.

Falciatore, A. and C. Bowler (2002). "Revealing the molecular secrets of marine diatoms." Annual Review of Plant Biology 53: 109-130.

Hagen, L., B. Kavli, et al. (2008). "Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA." The EMBO journal 27(1): 51-61.

Hazra, T. K., T. Izumi, et al. (2002). "Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA." Proceedings of the National Academy of Sciences of the United States of America 99(6): 3523-3528.

Hegde, M. L., C. A. Theriot, et al. (2008). "Physical and functional interaction between human oxidized base-specific DNA glycosylase NEIL1 and flap endonuclease 1." The Journal of Biological Chemistry 283(40): 27028-27037.

Hem, C. A. D. (2010). "Cloning, expression and characterizaton of an unusual DNA glycosylase in diatoms."

Imamura, K., A. Averill, et al. (2012). "Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA." The Journal of Biological Chemistry 287(6): 4288-4298.

Invitrogen "Instruction Manual: TOPO® TA Cloning® Kit."

Jacobs, A. L. and P. Schar (2012). "DNA glycosylases: in DNA repair and beyond." Chromosoma 121(1): 1-20.

Kavli, B., O. Sundheim, et al. (2002). "hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup." The Journal of Biological Chemistry 277(42): 39926-39936.

Krokan, H. E. and G. Slupphaug (1998). "[DNA repair enzymes and their genes]." Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke 118(13): 2037-2042.

Krokan, H. E., R. Standal, et al. (1997). "DNA glycosylases in the base excision repair of DNA." The Biochemical journal 325 (Pt 1): 1-16.

Kroth, P. G. (2007). "Genetic transformation: a tool to study protein targeting in diatoms." Methods in molecular biology 390: 257-267.

Lindahl, T. (1993). "Instability and decay of the primary structure of DNA." Nature 362(6422): 709-715.

Manser, L. (2002). GTPase Protocols The Ras Superfamily Human Press.

NanoDrop. "Instruction Manual: ND-1000 Spectrophotometer User's Manual."

Otterlei, M., E. Warbrick, et al. (1999). "Post-replicative base excision repair in replication foci." The EMBO journal 18(13): 3834-3844.

Pearl, L. H. (2000). "Structure and function in the uracil-DNA glycosylase superfamily." Mutation research 460(3-4): 165-181.

Poulsen Nicole, C. P., Kröger N. (2006). "Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae)." J Phycol(42): 470.

Prasad, R., D. D. Shock, et al. (2010). "Substrate channeling in mammalian base excision repair pathways: passing the baton." The Journal of Biological Chemistry 285(52): 40479-40488.

Reece, R. J. (2009). Analysis of Genes and Genomes, Wiley.

Siaut, M., M. Heijde, et al. (2007). "Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum." Gene 406(1-2): 23-35.

Slupphaug, G., I. Eftedal, et al. (1995). "Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase." Biochemistry 34(1): 128-138.

Stratagene. "Instruction Manual: QuikChange® Site-Directed Mutagenesis Kit." from http://kirschner.med.harvard.edu/files/protocols/Stratagene_quickchangepdf.pdf.

Theriot, C. A., M. L. Hegde, et al. (2010). "RPA physically interacts with the human DNA glycosylase NEIL1 to regulate excision of oxidative DNA base damage in primer-template structures." DNA Repair 9(6): 643-652.

Torseth, K., B. Doseth, et al. (2012). "The UNG2 Arg88Cys variant abrogates RPA-mediated recruitment of UNG2 to single-stranded DNA." DNA Repair.

Visnes, T., B. Doseth, et al. (2009). "Uracil in DNA and its processing by different DNA glycosylases." Philosophical transactions of the Royal Society of London. Series B, Biological sciences 364(1517): 563-568.

Wollen Steen, K., B. Doseth, et al. (2012). "mtSSB may sequester UNG1 at mitochondrial ssDNA and delay uracil processing until the dsDNA conformation is restored." DNA Repair 11(1): 82-91.

Zhao, X., N. Krishnamurthy, et al. (2010). "Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts." Biochemistry 49(8): 1658-1666.

Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annual review of genetics 43: 143-166.

Appendixes

1: Media and Solutions

2: Plasmid and vector maps

3: PCR primers

4: MALDI-TOF results

5: Calculation of specific UNG-activity

Appendix 1: Media and Solutions

LB medium (1L)	
Tryptone	10 g
Yeasy extract	5 g
NaCl	10 g

The reagents were dissolved and brought to 1 L with MilliQ water. The solution was autoclaved and allowed to cool to 50 °C before antibiotics were added. The solution was stored at room temperature.

LB agar plates (1L)	
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g

The reagents were dissolved and brought to 1 L with MilliQ water. The solution was autoclaved and allowed to cool to 50 °C before antibiotics were added.
The solution was poured into petri dishes, allowed to solidify and Stored at 4 °C.

2X YT medium (1L)	
Tryptone	16 g
Yeast extract	10 g
NaCl	5 g

The reagents were dissolved and brought to 1 L with MilliQ water. . The solution was autoclaved and allowed to cool to 50 °C before antibiotics were added. The solution was stored at room temperature.

f/2 agar plates (1L)	
Bacto agar	10 g
Filtrated and autocleaved sea water	500 ml
MilliQ water	490 ml
NaNO3 (stock)	1 ml
NaH2PO4 x H20 (stock)	1 ml
Trace metals (stock)	1 ml
Vitamins (stock)	0,5 ml
Zeocin (100mg/ml)	1 ml

Bacto agar, filtrated and autoclaved seawater and MilliQ water were mixed and autoclaved. The solution was allowed to cool before adding nutrition salts, antibiotics and vitamins. The solution was poured into petri dishes, allowed to solidify and Stored at 4 °C.

50X TAE buffer (1L)	
Tris base	242 g
Glacial acetic acid	57,1 ml
0,5 M EDTA pH 8	100 ml
MilliQ water	To final volume

The solution was mixed, autoclaved and stored at room temperature.

Stacking gel for SDS-PAGE	
MilliQ water	6,1 ml
0,5 M Tris-HCl, pH 6,8	2,5 ml
30% acrylamide	1,3 ml
10% SDS (w/v)	100 µl
10% APS (w/v) made fresh	50 µl
TEMED	10 µl

All reagents were mixed, except APS and TEMED. APS and TEMED were added to the solution right before use.

7% resolving gel for SDS-PAGE	
MilliQ water	15,3 ml
1,5 M Tris-HCl, pH 8,8	7,5 ml
30% acrylamide	6,9 ml
10% SDS (w/v)	300 µl
10% APS (w/v) made fresh	150 µl
TEMED	20 µl

All reagents were mixed, except APS and TEMED. APS and TEMED were added to the solution right before use.

10% resolving gel for SDS-PAGE	
MilliQ water	5,9 ml
1,5 M Tris-HCl, pH 8,8	3,8 ml
30% acrylamide	5 ml
10% SDS (w/v)	150 µl
10% APS (w/v) made fresh	150 µl
TEMED	6 µl

All reagents were mixed, except APS and TEMED. APS and TEMED were added to the solution right before use.

12% resolving gel for SDS-PAGE	
MilliQ water	3,4 ml
1,5 M Tris-HCl, pH 8,8	2,5 ml
30% acrylamide	4,0 ml
10% SDS (w/v)	100 µl
10% APS (w/v) made fresh	50 µl
TEMED	5 µl

All reagents were mixed, except APS and TEMED. APS and TEMED were added to the solution right before use.

5X SDS-PAGE gel electrophoresis buffer	
Tris base	0,5 M
Glycine	1,92 M
SDS	0,5% (w/v)
MilliQ water	To final volume

The solution was mixed and stored at room temperature.

Lysis buffer	
Tris-HCl pH 7/8,5	50 mM
NaCl	100 mM
Imidazole	10 mM
Glycerol	5% (v/v)
Triton X-100	0,2% (v/v)
DTT *	2 mM

*Added right before use.

1X complete EDTA-free protease inhibitor cocktail (Roche) was included in the solution (1 tablet per 100 ml). The solution was mixed and stored at 4 °C.

Histidine elution buffer 500 ml	
Tris-HCl pH 7/8,5	50 mM
NaCl	100 mM
Imidazole	500 mM
Glycerol	5% (v/v)
Triton X-100	0,2% (v/v)
DTT*	2 mM

*Added right before use.

The solution was mixed and stored at 4 °C.

GST elution buffer 500 ml	
Tris-HCl pH 7/8,5	50 mM
NaCl	100 mM
Imidazole	10 mM
Glycerol	5% (v/v)
Triton X-100	0,2% (v/v)
DTT*	2 mM
Glutathione*	10 mM

*Added right before use.

The solution was mixed and stored at 4 °C.

Dialysis buffer	
Tris-HCl pH 7/8,5	50 mM
NaCl	100 mM
DTT*	1 mM

*Added right before use.

The solution was mixed and stored at 4 °C.

5X SDS Sample buffer (10 ml)	
Tris 1M pH 6,8	2,5 ml
DTT	1,54 g
SDS	1 g
Bromophenol blue	0,5 g
Glycerol (85% Stock)	5 ml
MilliQ water	To final volume

Tris, DTT and SDS were mixed and incubated at 60 °C for 10 minutes, before the rest of the reagents were added. The buffer was brought to 10 ml with MilliQ water. The buffer was divided into aliquots and stored at – 20 °C.

f/2 medium			
Component	Stock solution (g/L dH_2O)	Quantity (ml)	Molecular concentration in final medium (M)
$NaNo_3$	75	1	$8,82 \times 10^{-4}$
$NaH_2PO_4 \times H_2O$	5	1	$3,62 \times 10^{-5}$
$NaSio_3 \times 9H_20$	45	1	$3,18 \times 10^{-4}$
Trace metal solution	See trace metal solution	0,5	-
Vitamin solution	See vitamin solution	1	-

950 ml filtrated natural seawater and 50 ml Milli-Q water was autocleaved. Reagents were added and the solution was sterile filtrated (0,20 µm). The solution was stored at 4 °C.

Trace metal solution			
Component	Stock solution (g/L dH_2O)	Quantity (ml)	Molecular concentration in final medium (M)
$CuSo_4 \times 5H_2O$	10	1	$4,2 \times 10^{-8}$
$ZnSO_4 \times 7H_2O$	22	1	$7,65 \times 10^{-8}$
$CoCl_2 \times 6H_2O$	10	1	$4,2 \times 10^{-8}$
$MnCl_2 \times 4H_2O$	180	1	$9,1 \times 10^{-7}$
$Na2MoO_4 \times 2H_2O$	3,3	1	$2,6 \times 10^{-8}$

Reagets were added to 950 ml MilliQ water. 5 g Fe-Sequestene (Fe-EDTA) was added and the final volume brought to 1 L with MilliQ water. The solution was stored at 4 °C.

Vitamin solution			
Component	Stock solution (g/L dH_2O)	Quantity	Molecular concentration in final medium (M)
Thiamine HCl	-	200 mg	$5,92 \times 10^{-7}$
Biotin	0,1	20 ml	$4,1 \times 10^{-9}$
Cyanocobalamin	1	1 ml	$7,38 \times 10^{-10}$

200 mg thiamine HCl was dissolved in 950 ml MilliQ water. Reagents were added and the final volume brought to 1 L with MilliQ water. The solution was sterile filtrated (0,2 µm) and stored at 4 °C.

Appendix 2: Plasmid and vector maps

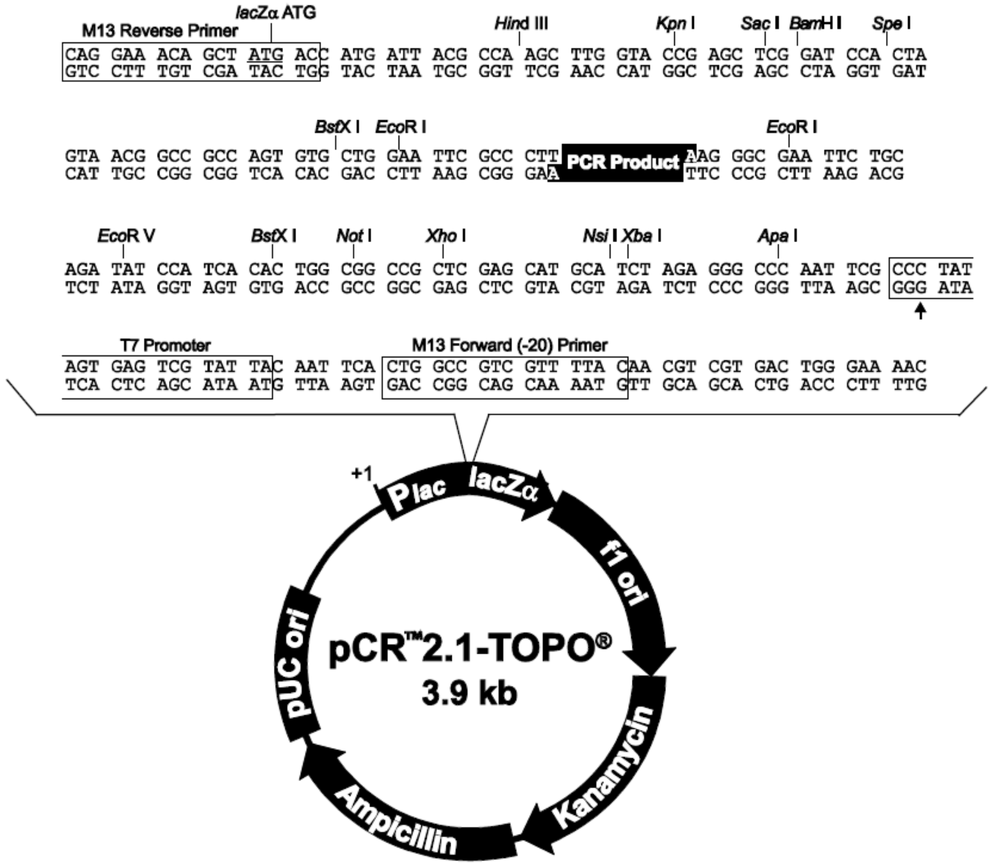


Figure 31: pCR™2.1-TOPO® plasmid map (Invitrogen)

Appendix 3: PCR primers

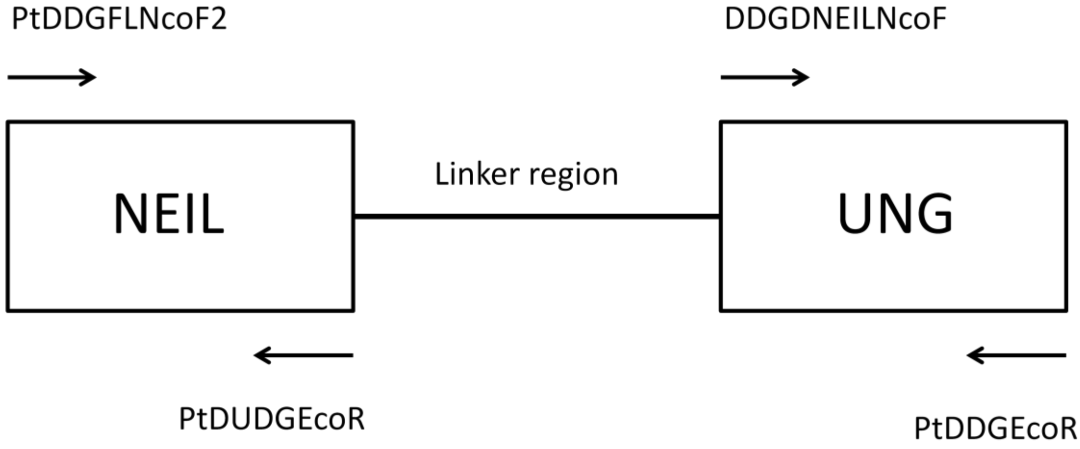


Figure 32: Schematic overview of the four primers used during PCR to produce PtDDG-FL, NEIL and UNG.

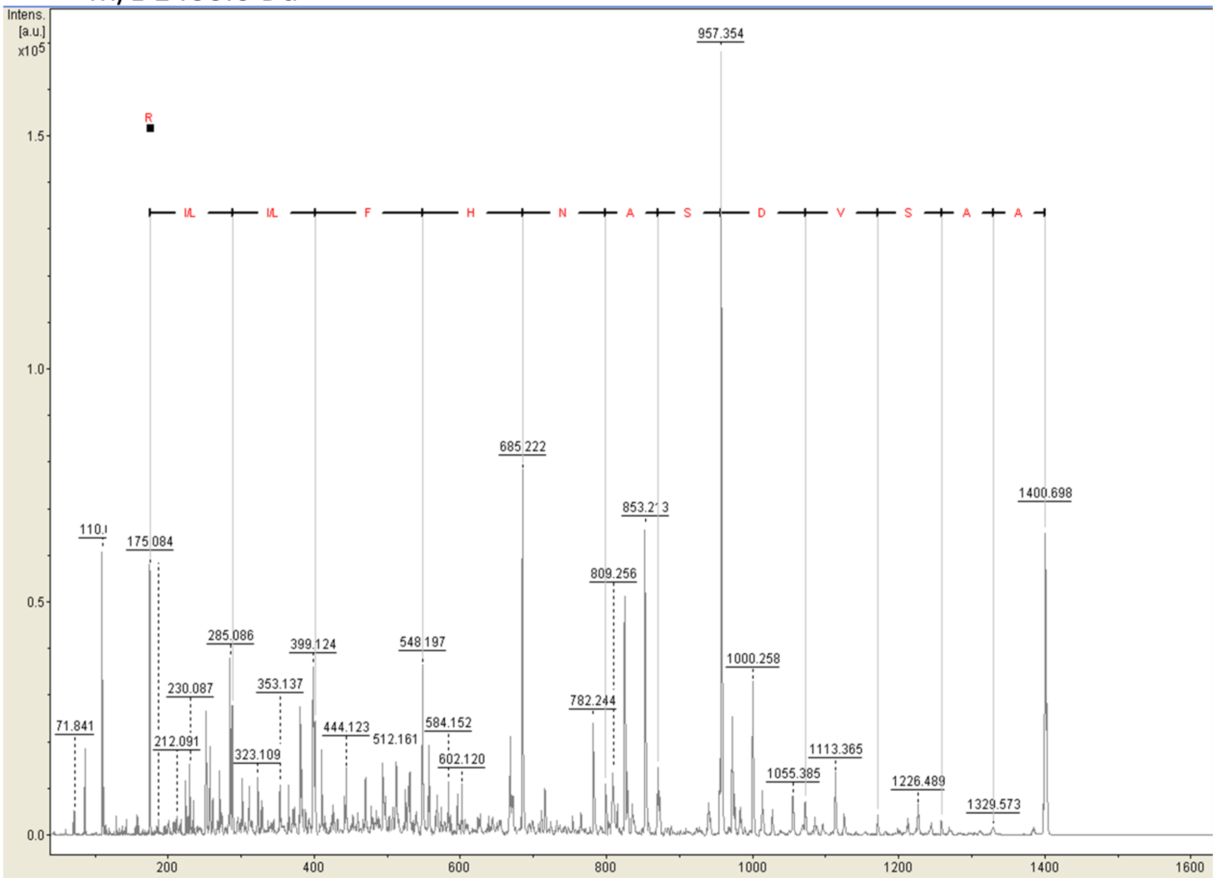
Primer	Primer sequence
DDGDNEILNcoF	AAGCCCATGGCCAGTTTGACAGAACA
PtDDGEcoR	TATGAATTCCTACTTGACGTTCCAATCGA
PtDDGFLNcoF2	CAGTCCATGGTGATGCCGGAAGGTCCAGAAG
DDGDUDGEcoR	CTGAATTCCTACACCCTCATCGCTCGTTC

PtDDG constructs	Primers used
NEIL	PtDDGFLNcoF2 and DDGDUDGEcoR
UNG	DDGDNEILNcoF and PtDDGEcoR
FL	PtDDGFLNcoF2 and PtDDGEcoR

Appendix 4: MALDI-TOF results

MSMS spectra
Sample: PtDDG-UNG
m/z 1400.6 Da

X-axis: mass (m/z)
Y-axis: intensity (a.u.)



Sample: PtDDG-UNG
m/z 1400.6Da

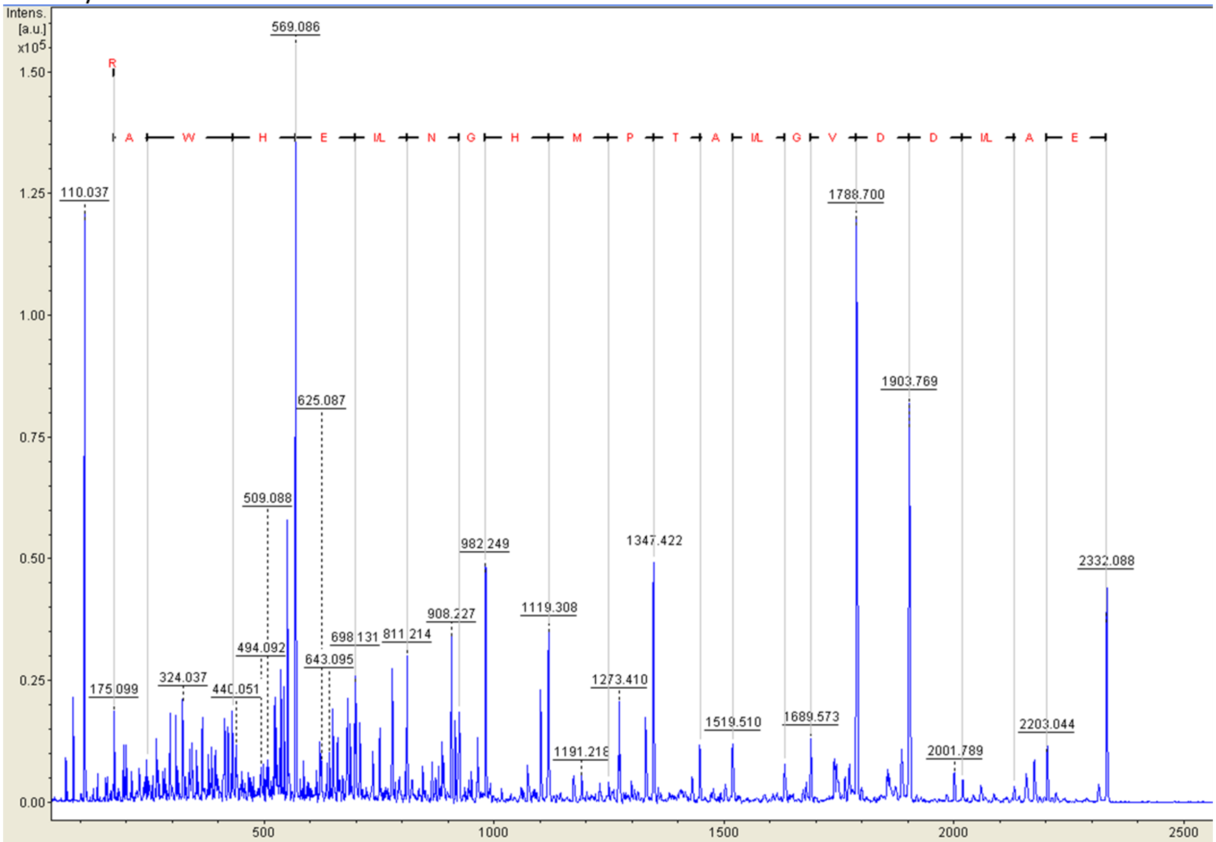
Match to: **gi|219118945** Score: **107 predicted protein [Phaeodactylum tricornutum CCAP 1055/1]**

Matched peptides shown in **Bold Red**

1 MASLTEQTWK EKLSTTFESE SFAKLAAFLE IERKMGATIY PPKEDIFSAL
51 NLCPFDKIKV VIVGQDPYHG PGQGHGLAFS VRRGVKPPPS LQNVFREAID
101 DVGIATPMHG NLEHWARQGV LLLNTVLTVR RGEANSHAGQ GWERVTDAII
151 HVVNAKSDGV VFLLWGNPAH KK**AASVDSAN HFIIR**SSHPS PLGATKTKSP
201 FLGSRCFSRA NDALKKMGKD PIDWNVK

MSMS spectra
Sample: PtDDG-UNG
m/z 2332.08 Da

X-axis: mass (m/z)
Y-axis: intensity (a.u.)



MSMS spectra
Sample: PtDDG-UNG
m/z 2332.08 Da

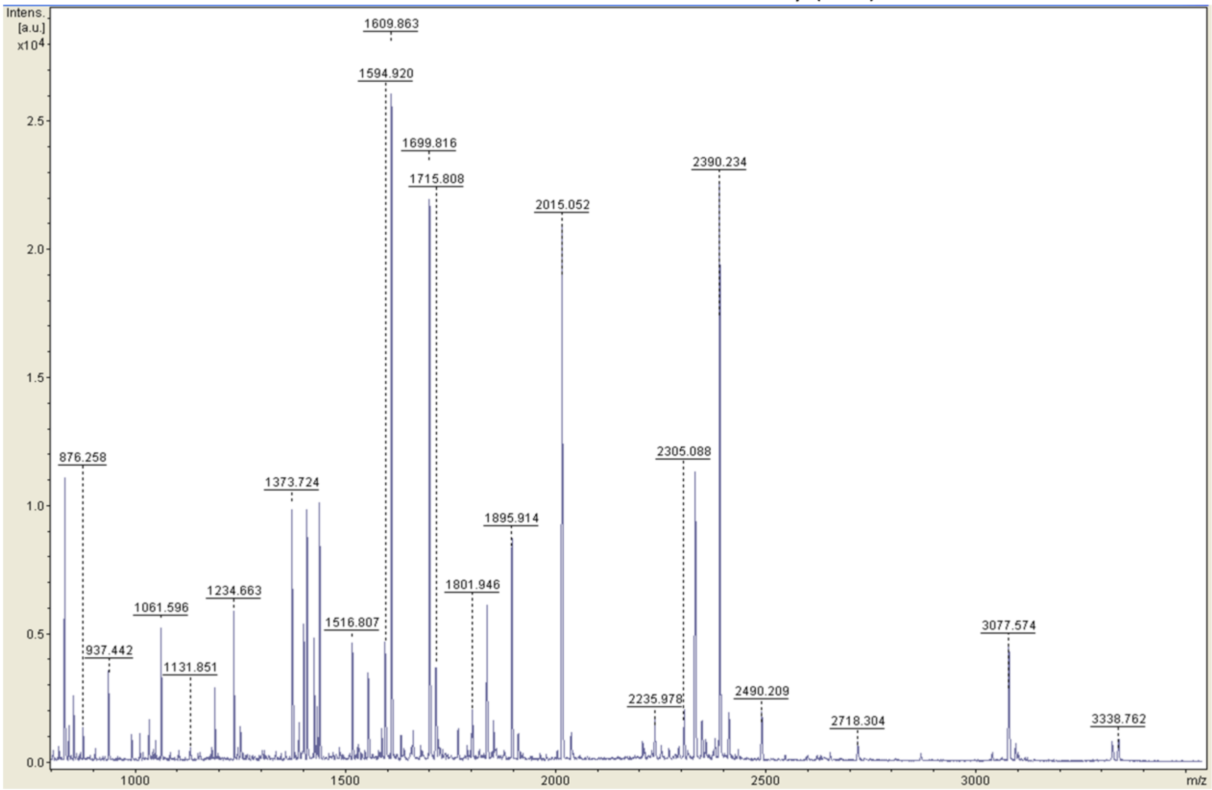
Match to: **gi|219118945** Score: **186 predicted protein [Phaeodactylum tricornutum CCAP 1055/1]**

Matched peptides shown in **Bold Red**

1 MASLTEQTWK EKLSTTFESE SFAKLAAFLE IERKMGATIY PPKEDIFSAL
51 NLCPFDKIKV VIVGQDPYHG PGQGHGLAFS VRRGVKPPPS LQNVFR**EAID**
101 **DVGIATPMHG NLEHWAR**QGV LLLNTVLTVR RGEANSHAGQ GWERVTDAII
151 HVVNAKSDGV VFLLWGNPAH KKAASVDSAN HFIIRSSHPS PLGATKTKSP
201 FLGSRCFSRA NDALKKMGKD PIDWNVK

MS spectra
Sample: PtDDG-FL

X-axis: mass (m/z)
Y-axis: intensity (a.u.)



MS spectra
Sample: PtDDG-FL

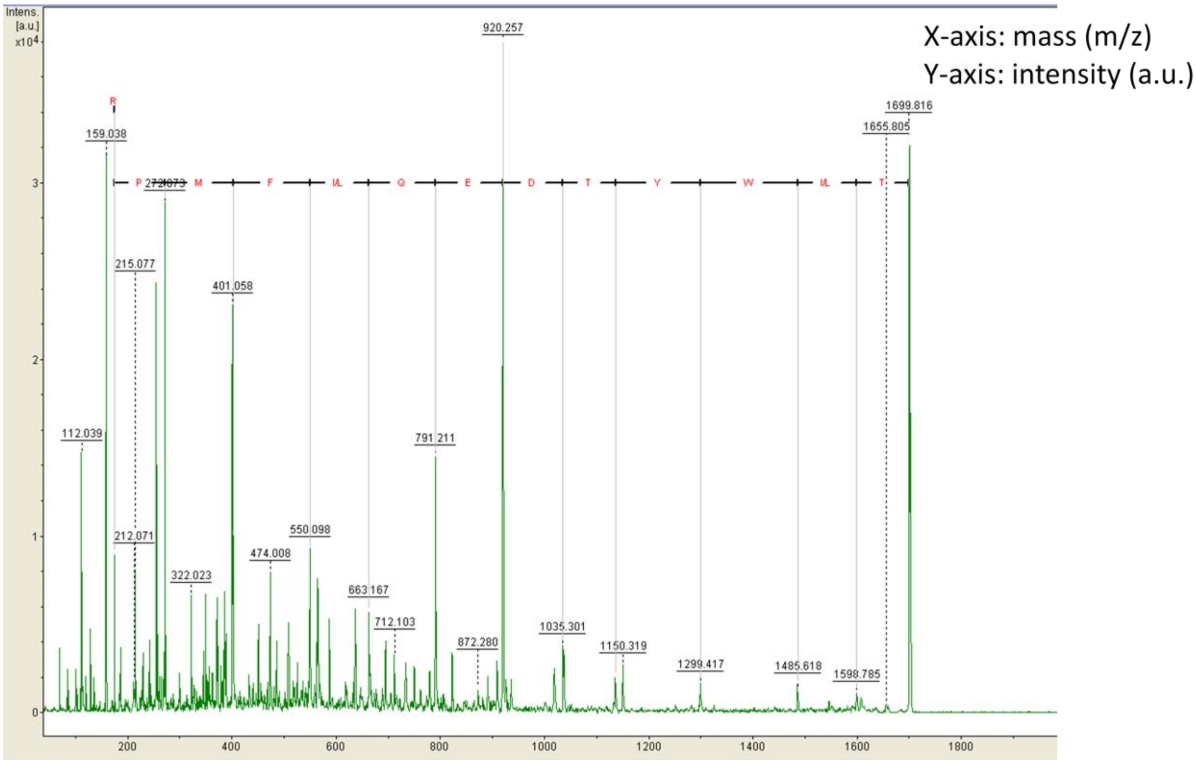
Match to: **gi|219118945** Score: **116** Expect: **2.6e-05 predicted protein [Phaeodactylum tricornutum CCAP 1055/1]**

Sequence Coverage: **65%**

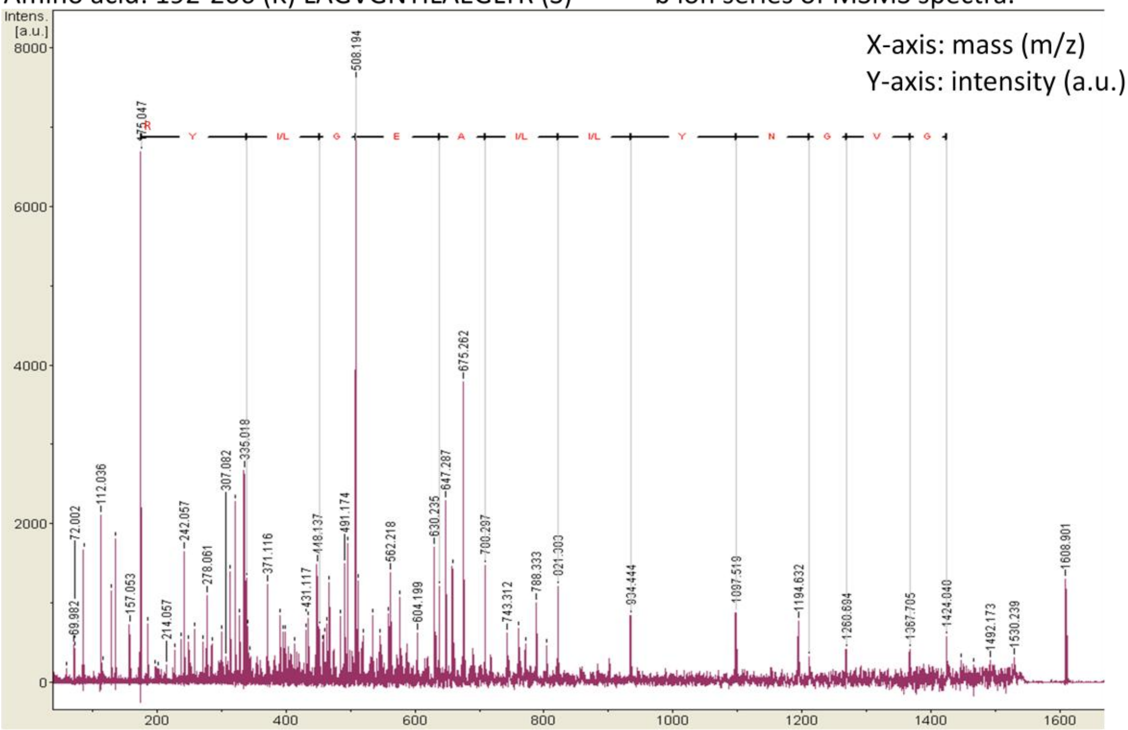
Matched peptides shown in **Bold Red**

1 MASLTEQTWK EKLSTTFESE SFAK**LAAFLE IERK**MGATIY PPK**EDIFSAL**
51 NLCPFDKIKV VIVGQDPYHG PGQGHGLAFS VRR**GVKPPPS LQNVFREAID**
101 DVGIATPMHG NLEHWARQGV LLLNTVLTVR RGEANSHAGQ GWERVTDAII
151 HVVNAK**SDGV VFLLWGNPAH KKAASVDSAN HFIIR**SSHPS PLGATK**TKSP**
201 FLGSRCFSRA NDALKKMGKD PIDWNVK

MSMS m/z 1699.8 Da   PtNEIL detected in PtDDG-FL sample
Sample: PtDDG-FL
Amino acid: 290-302 match in FL and NEIL amino acid sequence
(R) TIWYTDEQLFMPR (Q)



MSMS 1608.8 peptide PtDDG-NEIL
10 peptides matching the amino acid sequence in PtDDG-NEIL.
Sample: PtDDG-NEIL
Amino acid: 192-206 (K) LAGVGNYILAEGLYR (S)   b ion series of MSMS spectra.



Appendix 5: Calculation of specific UNG-activity

Calculation of specific UNG-activity

The specific activity is defined as nmol uracil cleaved from the substrate pr minute pr mg enzyme in the reaction. The two parameters needed are: amount of cleaved uracil and amount of protein in the sample.

Standard assay with [^{3}H] dUMP-labelled calf thymus-DNA

Amount of cleaved uracil: A β-counter are used to measure the sample dpm. The hot-to-cold ratio of radio-labeled dUMP in the substrate is known in forehand. From this ratio we know the specific activity of dUMP (mCi/µmol dUMP).

Calculation: 1dpm = 0,45 pCi

Specific substrate activity: 0,5 mCi/µmol dUMP

1 dpm = 0,9 fmol uncleaved dUMP (in total)

To calculate the amount nmol uracil cleaved per minute in a 20 µl reaction we use:

$$\frac{\dfrac{\text{Measured dpm}}{\text{minutes incubation (normalt 10)}} \text{ X } 0.9 \text{ fmol/dpm}}{10^6 \text{ fmol/nmol}} = \text{nmol dUMP cleaved}$$

Amount of protein in the sample:

$$\frac{\text{ng/}\mu\text{l protein x } 5\ \mu\text{l}}{10^6 \text{ ng/mg}} = \text{mg protein in the sample}$$

The specific activity calculation can be simplified to:

$$\frac{\textbf{Measured dpm x 0.018}}{\textbf{ng/}\boldsymbol{\mu}\textbf{l protein}} = \textbf{nmol dUMP/min/mg protein}$$