Identification and characterization of thermostable uracil glycosylases from the archaeon *Methanobacterium thermoautotrophicum* and the bacterium *Thermus thermophilus*

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Vytaute Starkuviene
aus Marijampole, Litauen

Göttingen 2001

Referent: **Prof. Dr. H.-J. Fritz**
Korreferent: **PD Dr. W. Kramer**

Tag der mündlichen Prüfung: 30 Oktober 2001

Angelei ir Juozapui

Table of contents

1. Introduction

1.1 Necessity of DNA repair

DNA is used as a carrier of genetic information in the majority of nowadays living organisms. The composition and structure of this molecule is well adapted to ensure the preservation of the coding information during the individual life-span and to enable it's accurate transfer from one generation to the next. Besides that, it maintains genetic variability, which is essential for adaptation and survival processes. However, DNA is a dynamic molecule and is subjected to constant changes. Many of them arise from natural cellular functions, like replication and recombination. Another source of modifications is chemical DNA instability under physiological conditions. In addition, DNA reacts easily with the various compounds. Each of these alterations affects the structure of the genetic material and is considered to be a DNA damage (Friedberg *et.al*., 1995). Vast majority of them changes the coding properties of DNA. Some modifications alter the efficiency of transcription, stall the replication fork, finally and lead the cells to apoptosis. Therefore, the existence of DNA damage recognition and removal processes is of crucial importance in the mutation avoidance and viability.

1.2 Sources of DNA damages and their repair mechanisms: overview

1.2.1 Repair of the DNA damages produced during replication

DNA replication is an essential cellular process, usually performed by a well regulated and multi-protein machinery. The fidelity of DNA synthesis is about one error per 10^{-9}-10^{-10} nucleotides (Kunkel, 1992) and is the sum of three factors: polymerase action, exonucleolytic proofreading activity and mismatch repair.

DNA polymerase is an extremely precise enzyme and synthesizes the DNA with only one mistake per 10^{-4}-10^{-5} nucleotides (Kunkel, 1992). This is achieved by a tight control on the stability of the newly made hydrogen bonds in the active center of the enzyme (Loeb and Kunkel, 1982). However, bias in the nucleotide pool, base tautomerisation, repetitive regions on the template leads to various misincorporations,

small insertions and deletions (Kronberg, 1992). The mistakes made by the DNA polymerase are efficiently removed by the 3'-5' exonucleolytic proof-reading activity (Goodman, 1988). The latter is usually associated with the replicative enzyme and contributes on average about 100-fold to the synthesis fidelity (Benkovic and Cameron, 1995). The replication errors, which escaped from the proof-reading, are repaired by the mismatch repair system MMR (Modrich and Lahue, 1997). This pathway is unique in that it recognizes normal, but not properly paired bases. Therefore, the restoration of the correct information requires distinguishing between the parental and newly synthesized strands. In difference to other repair machineries, MMR possess this feature. Thus, the action of the post-replicative MMR decreases the error rate in three additional orders of magnitude (Jiricny, 1998).

1.2.2 Repair of DNA damages produced between two rounds of replication

Vast majority of DNA damages occurs between the rounds of replication due to the interaction with the various endogenous and exogenous compounds. Consequently, cells have developed numerous and frequently overlapping defence mechanisms.

1.2.2.1 DNA damages of exogenous origin

Ionizing and UV (ultraviolet) radiation are considered the main sources of physical damages. The first produces variety of lethal lesions; the most dangerous of which are double stranded DNA breaks. Usually, they are subjected to repair by the recombination (Shinohara and Ogawa, 1995). UV light, though, is less energetic, but is absorbed by the DNA bases and forces covalent bond formation between the adjacent pyrimidines (Friedberg *et.al*., 1995). If not repaired, these bulky adducts create a block in the replication fork. Multi-enzyme nucleotide excision machinery (NER) acts on them by incising the DNA strand around the damage, followed by removal of the respective DNA fragment and, finally, the process is completed by the DNA re-synthesis step (de Laat *et.al*., 1999).

Among the most abundant exogenous reactants stand alkylating agents. Their interaction with DNA causes abnormal base methylation or AP-site formation. Some other chemicals, like nitrous acid, produce inter- and intrastrand DNA cross-links. Flat, multiple ring molecules are able to intercalate between the bases in the duplex

and serve as possible sites of the frameshifts in the following replication round (Friedberg *et.al*., 1995).

1.2.2.2 DNA damages of endogenous origin

DNA is mainly damaged by the naturally occurring hydrolysis, oxidation and non-enzymatic alkylation reactions. The latter frequently involves the action of S-adenosylmethionine (SAM), which transfers the alkyl group on the proteins and DNA (Lindahl, 1993). One of the products of DNA methylation, O^6-methylguanine pairs with thymine rather than cytosine, thus producing a point mutation. Such modifications are repaired by a simple damage reversal system. It involves the action of a single enzyme - guanine - O^6-specific methyltransferase, which removes the deleterious methyl group resulting in an error-free reconstitution of the normal guanine base (Moore *et.al*., 1994).

The most frequently DNA is damaged by oxidation reactions. Aerobically growing organisms are steadily exposed to reactive oxygen species, like hydroxyl and peroxide radicals or hydrogen peroxide (Friedberg *et.al*., 1995). As the major mutagenic lesion 7,8-dihydro-8-oxoguanine (8-oxoG) is formed. The latter pairs with adenine during the DNA replication (Shibutani *et.al*., 1991). Oxidation also damages pyrimidine bases. They form non-planar ring structures and, consequently, lose their coding properties (Breimer and Lindahl, 1984). Oxygen also attacks other cellular components, like unsaturated fatty acids. This generates by-products, which readily react with DNA molecule (Marnett and Plastaras, 2001). Typical end compounds of this pathway are exocyclic base modifications: etheno-A or etheno-C, which block normal hydrogen bond formation (Saparbaev and Laval, 1998). Various base alterations produced by oxidizing agents usually are the substrates of the base excision repair (BER) (more details in **1.4** section).

The third principal source of DNA damage is spontaneous hydrolysis reactions (Lindahl, 1993). Especially susceptible is the N-glycosidic bond of the purines (Lindahl and Nyberg, 1972). Depyrimidination, however, also occurs, but about 20 times slower (Lindahl, 1993). The resulting apurinic/apyrimidinic (AP) sites, if not repaired, form highly mutagenic DNA strand breaks. It was estimated that more than 9000 AP-sites are generated daily per human genome (Kunkel, 1999). Again, highly efficient BER repair (Doetch and Cunningham, 1990) reduce the number of AP-sites to 5-6 per genome.

Not only glycosidic bonds, but also DNA bases suffer from hydrolytic attacks. The exocyclic amino groups of the bases are labile and readily undergo reactions of hydrolytic deamination (Lindahl, 1993). Formation and repair of these DNA damages is the main focus of this work; therefore they will be described in more details.

1.3 Hydrolytic DNA deamination

1.3.1. Hydrolytic deamination of purines

In the course of the hydrolytic deamination, purines adenine and guanine are converted into the hypoxanthine and xanthine residues, respectively. Both modifications preferentially pair with cytosine (Karan and Lindahl, 1980). Thus, xanthine does not change coding properties of DNA, while hypoxanthine generates a pre-mutagenic lesion. But as rates of purines deamination are slow and the resulting products are repaired efficiently, no real threat to the integrity of the genetic information is considered.

1.3.2 Hydrolytic deamination of cytosine

A completely different picture evolves during the cytosine deamination process (Lindahl, 1993). The reaction occurs 40-50 times quicker than deamination of the purines (Lindahl and Nyberg, 1974). Resulting uracil is formed at especially high rates in the single stranded DNA during transcription, replication or recombination (Lindahl and Wood, 1999). Uracil is a highly mutagenic lesion due to its preferential pairing to adenine residue. Therefore, in following round of replication the transition mutation is fixed in half of the progeny. Besides hydrolysis, DNA replication is an additional source of uracil in the genome. Natural occurrence of dUTP's in the cell and non-discriminative action of the replicative DNA polymerase (Kronberg, 1992) enable missincorporation of uracil instead of thymine (Tye *et.al*., 1978). This, in general, does not change the coding information, but uracil-containing DNA possesses the altered binding affinities for the transcription factors or other regulatory proteins (Lindahl and Wood, 1999). About 400 uracil bases are generated daily by the hydrolytic deamination reaction in the human genome (Kunkel, 1999). Therefore, it represents one of the major threats to the genetic integrity. Uracil residues are

efficiently removed by a highly specific enzyme called uracil-DNA glycosylase (Lindahl, 1974), which initiates the cascade of BER events.

1.3.3 Hydrolytic deamination of 5-methylcytosine

Hydrolytic deamination occurs most rapidly at 5-methylcytosine sites (Lindahl, 1993). The estimated rates are about 4-5 times higher as for cytosine residues (Shen *et.al*., 1994). The resulting product is a natural thymine base, placed in opposition to the G (T/G mismatch). As in case of cytosine deamination, the subsequent replication rounds will generate a transition mutation.

Normally, 5-methylcytosine (5meC) is generated by the action of the specific DNA methyltransferases, which catalyze the transfer of a methyl-group from the S-adenosylmethionine (SAM) to the respective position on the cytosine ring (Vanyushin *et.al*., 1968). This type of DNA methylation is widely spread both in eukaryotes and prokaryotes (Rein *et.al*., 1998). The number of 5meC varies dramatically among different species: it hardly exceeds 1% of all cytosine bases in the prokaryotes (Vanyushin *et.al*., 1968), 0.1-0.3% in *Drosophila melanogaster* (Lyko *et.al*., 2000), 3-10% in mammalians (Frostesjo *et.al*., 1997), sometimes as high as 50% in the plants (Vanyushin and Kirnos, 1988). 5meC carries numerous important functions in the genomes. The prokaryotic microorganisms use this modification as a defense against the invasion of the foreign DNA species (Raleight, 1998), while in the eukaryotic cells 5meC known to be involved in the regulation of gene expression, embryogenesis, genomic imprinting, aging and some other, possibly still not elucidated processes (Grigg and Clark, 1994). Rapid deamination of the 5-meC makes the respective sites the mutational hotspots. Consequences of this, for example, humans are dramatic, like tumor formation, emergence of various genetic diseases and developmental abnormalities (Jones *et.al*., 1992).

In difference from uracil, the repair of T/G mismatches is intrinsically complicated and inefficient. The reason for this is, that not specifically altered, but normal, only mismatched bases needs to be recognized. The overall picture is further complicated by the only slight distortion in the duplex DNA around T/G mismatch (Hunter *et.al*., 1987). Nevertheless, during the past years, several highly specialized enzymes have been discovered. The majority of them are involved in BER pathway. In addition, few organisms utilize specific *very short patch repair* (VSP) mechanism (Lieb *et.al*., 1996).

1.4 Base excision repair

The base excision repair pathway apparently evolved to repair DNA damages, which cause only slight distortion of DNA molecule. BER is initiated by the action of the multiple DNA glycosylases, which recognize diverse base modifications. The action of these enzymes creates an AP-site. Then, an AP-site endonuclease cleaves the phosphodiester bond at 5' end of the AP-site. From this point, the BER branches into "short" and "long" patch repair. In case of first, the DNA polymerase β cleaves the AP-site at the 3' side by the fused activity of deoxyribophosphodiesterase (dRPase) and incorporates the missing nucleotide. Then, the appropriate ligase seals the gap. If the AP-site contains adducts, which are not substrates for the dRPase activity of polymerase β, the process is switched to the "long" patch. In this case, the concerted action of the PCNA (proliferating cell nuclear antigen), polymerases δ and ε, replication factor-C, displace the damaged strand by the synthesis of the longer fragment. The following actions of the flap-endonuclease and ligase complete the process. This scheme was reconstructed eukaryotic cells *in vitro*. It is believed, however, that BER in prokaryotes functions similar way, with the participation of the appropriate bacterial DNA polymerases (overview in Seeberg *et.al*., 1995; Parikh *et.al*., 1999).

1.4.1 DNA glycosylases

The main function of DNA glycosylases in BER is to hydrolyze the N-glycosidic bond between the base and deoxyribose. Mechanistically, the enzymes can be divided into two classes: monofunctional and bifunctional. The first group cleaves the glycosidic bond using water molecule as an attacking nucleophile and generates an AP-site. Glycosylases of the second group perform the attack by an internal amino residue. Then, base removal is followed by the β-elimination reaction, which produces break of phosphodiester bond at the 3' side of the AP-site (overview in David and Williams, 1998).

DNA glycosylases not only initiates the BER, but also comprise the damage-specific step, while following events of strand incision, DNA synthesis and ligation are damage-general. Several distinct DNA glycosylases recognize akylated bases. One of them, Tag protein, is fairly specific for the 3-methyladenine, while AlkA is induced upon the exposure to the alkylating agents and removes variety of damaged purines and pyrimidines (Bjelland and Seeberg, 1987). Several enzymes were also found for

repair of oxidized bases. Oxidized pyrimidine derivatives are mainly removed by the Nth (endonuclease III) (Breimer and Lindahl, 1984). Oxidized purines, like 8-oxoG, are the substrates of the MutM enzyme (Nash *et.al*., 1996). Adenine, missincorporated opposite to the 8-oxoG, is removed by the MutY glycosylase (Michaels *et.al*., 1992). Some of the organisms, like bacteriophage T4, *Micrococcus luteus* possess glycosylases, which are able to remove pyrimidine dimers produced by the UV irradiation (overview in David and Williams, 1998).

Remarkably, activities, specific for the removal of damages of hydrolytic base deamination, comprise the broadest group (Table 1). Since these enzymes are the main focus of this work, their properties, distribution and ways of discovery are presented in details in the following sections.

1.4.1.1 Glycosylases, specific for the repair of the damages caused by the hydrolytic deamination

The first enzyme specific for the removal of uracil was discovered in *E.coli* cells and named UDG - uracil-DNA glycosylase (Lindahl, 1974). Subsequently, similar enzymes were found in different organisms, varying from viruses to the higher eucaryotes. UDG is highly conserved, and strictly uracil specific enzyme. It removes the latter from U/G mismatches and U/A matches, double and single stranded DNA (overview in Krokan *et.al*., 1997). Therefore, it is believed that UDG is the main activity to protect the genome from the mutagenic uracil residue. The structural basis of UDG specificity and mechanism of action are well understood due to the availability of the three dimensional structure of the enzyme (Mol *et.al*., 1995). Highly selective pocket disfavors binding of any other base. Uracil can be accommodated only in the flipped out conformation, which facilitates the access to the glycosidic bond (Pearl, 2000).

Another uracil glycosylase of very general action was discovered almost three decades later. First representative was isolated from the thermostable bacteria *Thermotoga maritima* (TM) by Sandigursky and Franklin, 1999. As the enzyme showed similar biochemical properties to UDG, it was named TMUDG glycosylase. However, no significant similarity on the level of primary protein composition was detected. With the accumulation of the genomic data, it became obvious that the enzyme is also widely spread, however, only within the microbial world. No eucaryotic representative was found to date (unpublished observations). Unfortunately, the investigation of these enzymes is still in the initial phase, thus, the exact biological role and mechanism of action awaits to be clarified.

Base excision repair (BER)

Subfamily	Documented occurrence	First biochemical description
UDG	Everywhere, except archaea: *E.coli, S.cerevisiae, H.sapiens*	Lindahl, 1974.
Mig	Archaea: *Methanobacterium thermoautotrophicum*	Horst and Fritz, 1996.
TDG/MUG	Eukaryotes, bacteria*: H.sapiens, E.coli*	Neddermann *et.al*., 1996., Gallinary *et.al*., 1996.
MBD4	Eukaryotes: *H.sapiens*	Hendrich *et.al*., 1999.
hSMUG	Eukaryotes: *H.sapiens, X. laevis*	Haushalter *et.al*., 1999.
TMUDG	Everywhere, except eukaryotes: *T.maritima, A.fulgidus*.	Sandigursky and Franklin, 1999

Very short patch repair (VSP)

Subfamily	Documented occurrence	First biochemical description
Vsr	Bacteria: *E.coli*	Hennecke *et.al*., 1991

Table 1. List of known activities involved in the repair of damages of hydrolytic deamination.

Recently, a new eucaryotic uracil glycosylase hSMUG was isolated (Haushalter *et.al*., 1999). Seemingly, it could be eukaryotic specific enzyme, while only the representatives of latter kingdom contain the respective genes. Initial measurements indicated that the enzyme is active on the single stranded DNA. Later, it was shown that the interaction with the AP-site endonuclease significantly stimulated the activity also on the double stranded DNA (overview in Schärer and Jiricny, 2001). Thus, very likely, the enzyme performs the back-up function for UDG in the genome. hSMUG also has a unique sequence.

Several enzymes were found to be specific to T/G mismatch. First isolated was human TDG glycosylase (Nedderman and Jiricny, 1993). Subsequently, it was shown that the enzyme also process U/G efficiently. No cleavage of single stranded DNA or U/A match was observed. Shortly after, bacterial homolog was isolated and named MUG (Gallinary and Jiricny, 1996). Crystal structure of the latter (Barret *et.al*., 1998) revealed the surprising similarity to UDG enzymes, despite negligible similarity of their protein sequences. Three-dimensional structure also clarified the mode of action of the TDG/MUG enzymes. The substrate specificity is determined not only by the active pocket of the enzyme, but also by the nature of the opposite

base. Preferential cleavage of T/G mismatch embedded into the CpG sequence context indicates that the repair of thymine residues arising from the deamination of 5meC in the CpG sequences is the main role of the enzyme (Waters and Swann, 2000).

Few years later, T/G, U/G mismatch specific Mig.*Mth* glycosylase was isolated from the thermostable archaea *Methanobacterium thermoautotrophicum* THF (Horst and Fritz, 1996). Mig.*Mth* shares the same substrate recognition properties as TDG. However, no sequence homology was detected. Mig gene is located in the close proximity to the gene of cytosine specific methyltransferase (Vos and Nölling, 1992). Thus, the functional coupling of the activities of both proteins was suggested.

One more T/G, U/G specific glycosylase MBD4 was recently isolated from human cells (Hendrich *et.al*., 1999). The enzyme is similar to the Mig glycosylase according the biochemical properties and amino acid sequences. However, it is unusual in that it is not a stand-alone protein, but fused to the CpG binding domain. Thus, as in the case of Mig and TDG enzymes, MBD4, most likely, is involved in the repair of the products of 5-meC deamination within the particular sequence context. Removal of uracil occurs, most likely, due to the structural similarity and could serve as a back-up function to the "true" uracil-glycosylases.

Finally, T/G and U/G mismatches are repaired not only by BER pathway, but also by VSP. The attack on the T/G mismatch is initiated by the Vsr endonuclease (Hennecke *et.al*., 1991), which incises DNA at the 5' side of the T/G mismatch. Then, coordinated action of the exonuclease and/or polymerase restores the genetic information. In addition to that, interaction with the MutS and MutL proteins occurs, which increases the affinity of the Vsr to T/G mismatch (Drotschmann *et.al*., 1998). The gene of the vsr overlaps with the gene of cytosine specific dcm methyltransferase (Sohail *et.al*., 1990). Sequentially, functional interaction of Vsr and Dcm proteins is well-documented (Gläsner *et.al*., 1995). However, in difference from the BER, the distribution of this repair system is very limited. The representatives were isolated only from the few bacteria, like *E. coli* (Hennecke *et.al*., 1991) and *B. stearothermophilus* (Laging, 2000).

1.5 Discovery of the activities specific for the repair of the damages caused by hydrolytic deamination

The list of the glycosylases specific for damages caused by hydrolytic deamination is constantly growing (Table 1). Consequently, variety of approaches has been implemented to facilitate detection and isolation of the new enzymes.

1.5.1 Biochemical methods

Investigation of the hydrolytic DNA degradation in the aqueous solutions raised the suggestion that cells may contain the activity responsible for the repair of the uracil residues (Lindahl and Nyberg, 1974). Thus, the search for such activities was undertaken (Lindahl, 1974). Using a radiolabelled dU substrate, the uracil-releasing activity was isolated from the cells extracts of *E. coli* by the sequential purification steps. This enzyme was UDG and was the first glycosylase isolated. Similar approach also worked well for the discovery of T/G-mismatch specific enzyme TDG (Neddermann and Jiricny, 1993).

1.5.2 Genetic methods

The discovery of Vsr endonuclease was based on genetic methods. Shortly after characterization of the cytosine specific Dcm methyltransferase in *E. coli*, the presence of the very short patch repair (VSP) was described (Lieb *et.al*., 1986). Early works on VSP revealed, that some of the dcm gene mutants led to loss of this pathway. Therefore, the conclusion was made that Dcm protein is directly involved in this process. Separation of the Dcm and VSP phenotypes came with the construction of the plasmid, carrying fragment of the *E. coli* chromosome and complementing both activities. Library of sequential deletions of this plasmid resulted in the dcm^- vsp^+ phenotype. Sequencing of the respective insert revealed the presence of the unknown *orf* vsr (Sohail *et.al*., 1990). Following works characterized Vsr as strand and sequence specific DNA endonuclease (Hennecke *et.al*., 1991).

1.5.3 Proteomics

Traditional biochemical and genetic methods have been quite successful in identifying some of the DNA repair proteins. However, the isolation of the novel

activities requires broader screening strategies. The advent of large-scale protein analysis – proteomics seems to be particularly promising.

One of the techniques, named expression cloning, enables to detect the specific enzymatic or binding activity without *a priori* knowledge of the gene or the purification of the gene product (Lustig *et.al*., 1997). For this, the cDNA library of interest is sub-cloned, and following *in vitro* steps of transcription and translation results in a pool of proteins. The latter can be subjected to the desired screening. Haushalter *et.al*., 1999, applied the mixture of the synthetic analogs of the AP-site, which act as glycosylase inhibitors. The fraction of the proteins with positive signal was obtained and refined to a single cDNA clone. The latter was sequenced and proved to code for the novel type of uracil-glycosylase, named hSMUG.

1.5.4 Genomics

The progress in automated sequencing techniques leads to the accumulation of vast amounts of data. Their analysis by computational techniques offers numerous possibilities. Similarity searches performed against conceptual translations of sequences with the protein of interest can identify genes encoding related proteins. Sequentially, comprehensive phylogenetic analysis and reconstruction of the metabolic pathways can be performed (Aravind *et.al*., 1999). Searches based on the conserved motives are invaluable in finding the genes with a high degree of divergence Application of more sophisticated working tools helps to assign a precise biological function within the particular protein family (Gogos *et.al*., 2000).

Homology searches were used for the identification of the novel uracil glycosylase TMUDG (Sandigursky and Franklin, 1999). The protein was discovered due to the weak similarity to MUG glycosylase. Despite that the Blast search output was statistically weak, it was falling into the region of the conserved catalytic motif, which is essential for the hydrolysis of glycosidic bond. Enzymatic characterization of the respective *orf* revealed new family of uracil glycosylases.

1.6 The aim and methodology of the work

Hydrolytic attack on DNA produces a variety of alterations, number of which increases significantly at high temperatures (Lindahl and Nyberg, 1972). The rates of hydrolytic deamination are particularly astonishing: cytosine conversion to the uracil at 95 ^{o}C increases more than 1000 folds (Lindahl, 1993). Thus, the discovery of

thermophilic microorganisms (Brock and Darland, 1970) raised a lot of interest in understanding how their genomic information remains intact. Despite some progress in this field, the list of the strategies to counteract the damages produced by the rapid hydrolytic deamination at high temperatures is far from being completed. Not much is known about the interplay of the repair with the other cellular processes, like replication and transcription. Therefore, the studies of the maintenance of the genetic fidelity in thermophiles are exciting and challenging topic.

In this work the repair of damages caused by hydrolytic DNA deamination were investigated in several microorganisms. One of them is a moderate thermophile archaea *Methanobacterium thermoautotrophicum* ΔH. The choice for it was made due to several reasons. First, it is one of the earliest completely sequenced organism. Second, the original annotation failed to find any U/G, T/G - specific activities (Smith *et.al.*, 1997). Another thermophile of choice is bacteria *Thermus thermophilus* HB27. As the genome of this microorganism is currently sequenced in the Göttingen Genomics Laboratory, Institute für Mikrobiologie und Genetik, Georg-August Universität Göttingen, not-published data are available for the first-hand analysis. *T. thermophilus* is of the particular interest due to its high growth temperatures (about 85^{o}C) and richness in GC bases (about 70%). Thus, highly increased rates of uracil formation are expected. Besides that, *T. thermophilus* possess transformation system, that makes it suitable for the genetic studies and direct measurement of the impact of various repair systems onto the genetic stability.

As a first phase of work, the computational searches of the U/G, T/G specific glycosylases and/or endonucleases were defined. Then, isolation of *orf*s candidates and biochemical characterisation of the putative DNA repair proteins follows. The latter knowledge combined with the analysis of the surrounding genomic locus serves for the refinement of biological role of respective enzymes.

In case of *M. thermoautotrophicum* ΔH, particular attention was directed to the *orf* 496. Already initial sequence comparisons revealed that *orf* 496 is a close homolog of T/G, U/G mismatch glycosylase Mig.*Mth* (Horst and Fritz, 1996). The latter was found on the plasmid pFV1 of *M. thermoautotrophicum* THF strain, in the close proximity to the cytosine-specific methyltransferase (Nölling *et.al.*, 1992). Therefore, functional interaction of methyltransferase and Mig.*Mth* glycosylase was suggested (Horst, 1996). However, detailed kinetic assays revealed that the enzyme is only little discriminative towards sequence context (Fondufe, 1999). Accordingly, Mig.*Mth* glycosylase could serve as a general defence mechanism against deamination damages in *M. thermoautotrophicum*. Doubts, however, are raised by

plasmid localization of Mig.*Mth.* In difference to that, *orf* 496 is located on the chromosome of the *M. thermoautotrophicum* ΔH. Therefore, enzymatic characterisation of the putative protein encoded by the *orf* 496 is highly advantageous in clarifying exact biological role of Mig glycosylases.

Work devoted to *T. thermophilus* concentrates on the characterization of two *orf*s, which have high degree of homology to the recently discovered TMUDG glycosylase (Sandigursky and Franklin, 1999). Despite, that redundancy of uracil removing activities is well spread in prokaryotic world; this usually means the presence of glycosylases and/or endonucleases of different families. Co-localisation of two genes, belonging to the same family is found for the first time. Therefore, the analysis of both putative proteins, named TTUDGA and TTUDGB is highly challenging.

2. Materials

2.1 Strains

The strains of microorganisms are presented in the following table:

Escherichia coli	**Genotype**	**Reference**
BL21(DE3) pLysS	*E.coli*B F$^-$ *dcm ompT hsdS*($r_B^-m_B^-$) *gal* λ(DE3) pLysS (*Cm*^{r})	Studier and Moffat, 1986
BL21(DE3)	*E.coli*B F$^-$ *dcm ompT hsdS*($r_B^-m_B^-$) *gal* λ(DE3)	Studier and Moffat, 1986
DH5α	*supE*44 Δ*lac*U169 (φ80*lacZ*ΔM15*)* *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1	Hanahan, 1983
TOP10 One Shot™	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) (φ80*lacZ*ΔM15) Δ*lac*X74*rec*A1 *deo*R *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^r) *end*A1 *nup*G	Invitrogen, Groningen
BW310	*ung spoT*1 *thi*-1 *relA*1 λ$^-$	Duncan and Weiss, 1982
BW313	*dut ung spoT*1 *thi*-1 *relA*1 λ$^-$	Duncan and Weiss, 1982
CJ236	*dut ung spoT*1 *thi*-1 *relA*1 λ$^-$ pCJ105(*Cm*^{r})	Joyce and Grindley, 1984
K38	HfrC (λ)	Russel and Model, 1984
Methanobacterium thermoautotrophicum	**Genotype**	**Reference**
DeltaH		Zeikus and Wolfe, 1972

Table 2. List of the different strains of the microorganisms used in this work and their genotypes, if available.

BL21 is an *E.coli*B strain lacking *vsr* gene. This strain also does not has *lon* ATP-dependant protease and *ompT* outer membrane protease, what decreases the degradation of recombinant proteins during purification process. BL21 carries DE3 lysogen, coding the gene of the T7 RNA polymerase, thus are used as an expression host for the proteins regulated by the T7 promotor. BL21(DE3)pLysS carries pLysS plasmid, which encodes the T7 lysocyme. The T7 lysozyme is a natural inhibitor of the T7 RNA polymerase and, consequently, provides additional stability to the expression of the target genes. Due to its ability to cut the peptidoglycan layer of *E.coli*, the T7 lysozyme is an efficient mean to induce the rapid lyses of the cells (Novagen, Madison). Finally, Bl21 strains are *dcm* deficient and can be used for the expression and enzymatic characterisation of the 5-cytosine specific methyltransferases.

DH5α were used for the cloning/recloning experiments due to the deficiency in the recombination process.

Topo10 OneShot™ cells are also deficient in the recombination. It was supplied as a chemically competent strain by Invitrogen and used for the cloning experiments with the topoisomeraseI.

BW310, BW313 and CJ236 cells were used for the expression experiments with the uracil-glycosylases from *Thermus thermophilus*. All these strains are deficient in the ung glycosylase gene. BW313 and CJ236 are also lacking the dUTP-ase gene (dut) and, consequently, contain high amount of uracil residues in their genomic DNA. The strain CJ236 is produced by mating F-factor pCJ105 into BW313 strain.

K38 was used as a host for the production of pGP1-2 plasmid.

2.2 Media

2.2.1 Media for *Escherichia coli*

2 x YT medium

16 g Bacto-trypton, 10 g yeast extract, 5 g NaCl, add 1l H_2O and autoclave (Sambrook *et.al*., 1989).

LB medium

10 g Bacto-trypton, 5 g yeast extract, 10 g NaCl, add 1l H_2O and autoclave (Sambrook *et.al*., 1989).

LB agar

10 g Bacto-trypton, 5 g yeast extract, 10 g NaCl, 15 g agar, add 1l H_2O and autoclave (Sambrook *et.al*., 1989).

SOC medium

20 g Bacto-trypton, 5 g yeast extract, 0.5 g NaCl, 20 ml 1M glucose, 10 ml 250 mM KCl, 5 ml 2M $MgCl_2$, add H_2O till 1l and autoclave (Sambrook *et.al*., 1989).

For the preparation of the selective media, sterile solutions of antibiotics are added to the medium after autoclave. The end concentrations were as follows:

Ampicillin:	50 µg/ml (pET21d-vector) (Novagen, Madison)
Kanamycin:	50 µg/ml (TOPO^{TM} cloning kit) (Invitrogen, Groningen)
	30 µg/ml (pGP1-2 plasmid) (Tabor and Richardson, 1985)
Chloramphenicol:	34 µg/ml (pLysS-vector) (Novagen, Madison)
	15 µg/ml (pCJ105-episome) (Joyce and Grindley, 1984)

Lac-promoter was induced with IPTG with the end concentration of 1 mM.

2.2.2 Medium for *Methanobacterium thermoautotrophicum* ΔH

0.3 g KH_2PO_4, 1.5 g $(NH_4)_2$ SO_4, 0.6 g NaCl, 0.12 g $MgSO_4$ x $7H_2O$, 0.08 g $CaCl_2$ x 2 H_2O, 4.0 mg $FeSO_4$ x $7H_2O$, 0.15 g K_2HPO_4, 4 g Na_2CO_3, 10 ml vitamin solution, 10 ml trace element solution, 1.0 mg resazurin, 1.5 g cysteine-HCl x H_2O, 1.5 g Na_2S x $9H_2O$, add 1l H_2O, pH 7.2 (Nölling *et.al*., 1991).

Vitamin solution

2 mg biotin, 2 mg folic acid, 10 mg pyridoxine-HCl, 5 mg thiamine-HCl, 5 mg riboflavine, 5 mg nicotinic acid, 5 mg Ca-pantothenate, 0.1 mg vitamin B_{12}, 5 mg p-aminobenzoic acid, 5 mg α-lipoic acid, add 1l H_2O. Kept at 4 ^{o}C, in the dark (Wolin *et.al*., 1964).

Trace elements solution

100 mg ZnSO4, 30 mg MnCl2 x $4H_2O$, 300 mg H3BO3, 200 mg CoCl2 x $6H_2O$, 10 mg CuCl2 x $2H_2O$, 20 mg NiCl2 x $6H_2O$, 30 mg Na2MoO4, add 1l H_2O. (Pfenning and Lippert, 1966).

2.3 Plasmids

Plasmids used in this work are listed in the **Table 2**. Reference indicates the source of information where details of the plasmid organization can be found.

Vector	Application	Reference
pET-21d	Vector used for the expression of all proteins described in this work under the control of the bacteriophage T7 promotor.	http://www.novagen.com/html/vectfram.html
pGP1-2	As a source of the T7 RNA polymerase for the protein expression in the *E.coli* strains lacking the respective gene.	Tabor and Richardson, 1985
pCR-Blunt II-TOPO	Vector for the cloning blunt-end PCR products by the activity of the topoisomerase I	http://www.invitrogen.com/vectordata/index.html
pCR 2.1-TOPO	Vector for TA cloning of PCR products by the activity of the topoisomerase I	http://www.invitrogen.com/vectordata/index.html

Table 3. Vectors used in this work.

2.4 Oligodesoxyribonucleotides

2.4.1 Oligonucleotides for sequencing

Name and length of the primer	Sequence	Application
T7 promoter primer (22mer)	5'-TTAATACGACTCACTATAGGGG	For the pET-21d vector
T7 terminator primer (19mer)	5'-GCTAGTTATTGCTCAGCGG	For the pET-21d vector

M31 Forward (24mer)	5'-ACGACGTTGTAAAACGACGGCCAG	For the TOPO™ vectors
M13 Reverse (24mer)	5'-TTCACACAGGAAACAGCTATGACC	For the TOPO™ vectors
seqMTHconf (20mer)	5'-GCCCGTTCTCTGACCGTAAG	For the region at 5' side of *mig2*
seqMTHmid (19mer)	5'-GTTTCCTGTCCCGCTTCAC	Internal primer of the *mig2* gene
MTHmetmidb (19mer)	5'-GCGCGAAGACACAACCCGC	Walk on M.*Mth*H methyltransferase gene

2.4.2 Oligonucleotides for DNA amplification

Underlined nucleotides indicate the restriction sites, when incorporated.

Name and length of the primer	Sequence	Product
Ncochrom (27mer)	5'-ATCTACAGGGGGGACCCATGGTTTCTG	Full-length *mig2* gene
496Nco (31mer)	5'-GAATTCTGCTGGCTGAAACCATGGTTCACAG	Truncated *mig2* gene
496Xho (29mer)	5'-CATTGTTTTTCTCATTCTCGAGTCTCCTG	Truncated and full-length *mig2*
MTHconf (28mer)	5'-GAGAACAGGCCCGTTCTCTGACCGTAAG	To the 5' side extended *mig2*
Ncometilaze (30mer)	5'-GAGGCAGGTTCGGCGATGTCAGATGTTGTG	M.*Mth* without restriction site
NcometilazeI (30mer)	5'-GAGGCAGGTTCGGCCATGGCAGATGTTGTG	M.*Mth* with restriction site
Xhometilaze (31mer)	5'-CCTATTCAGAAACTCGAGGTCCCCCCTGTAG	M.*Mth* with restriction site
thermusNcob (31mer)	5'-CAGCTCTAGCCATGGAGGCCTGGCAGAAAGC	MutY.*Tth* with restriction site

thermusHind (28mer)	5’-ACAGTGCAAGCTTTGCGTCCGGGAGGGG	MutY.*Tth* with restriction site
thermusN (24mer)	5’-GTGGAGGCCTGGCAGAAAGCCCTC	MutY.*Tth* without restriction site
thermusC (24mer)	5’-CTATGCGTCCGGGAGGGGGACTAC	MutY.*Tth* without restriction site
X216Nco (20mer)	5’-AAGGAGGGGCCATGGAGTTC	X216 with restriction site
X216Xho (28mer)	5’-GCAGTTCCAGGGTCACTCGAGGGGCTTG	X216 with restriction site
TTUDGANco (30mer)	5’-CCGCAAGCCCCTGCCATGGCCCTGGAACTG	TTUDGA with restriction site
TTUDGAXho (29mer)	5’-CGCGGGGGCTTACTCGAGGGGCTCCTGGC	TTUDGA with restriction site
TTUGDBNco (31mer)	5’-CGACAACATCCCCCTGCCATGGACAGGGAAG	TTUDGB with restriction site
TTUGDBEag (44mer)	5’-TGGACTACGAGGACCTCCTCTCCCGGCCGAAAGCCCGGCGAGGC	TTUDGB with restriction site
TTUGDBupper (31mer)	5’-CGACAACATCCCCCTGTGATGGACAGGGAAG	TTUDGB without restriction site
TTUGDBlower (44mer)	5’-TGGACTACGAGGACCTCCTCTCCTGGCTCAAAGCCCGGCGAGGC	TTUDGB without restriction site
TTUGDBfar (26mer)	5’-AGCTCGCCGTGGGCACGGCGCAACGG	TTUDGB extended to 3’ end
PolNot (40mer)	5’-CCTGGAGGTCCTGATGGAGGGCGGCCGCCTCGCCGGGCTT	PolX.*Tth* with restriction site
PolNhe (32mer)	5’-GTATAGTGGGGGTATGGCTAGCCAGGAGCTTG	PolX.*Tth* with restriction site

2.4.3 Oligonucleotides for the measurements of enzymatic activity

Some of the oligos used were taken from Dr. Yvonne Fondufe-Mittendorf and Dr. Martin Laging work, with the names left the same. Bold letters indicates variable sequence, underlined bases is mismatched, (F) – fluorescein attached to the 5' end of oligonucleotide.

2.4.3.1 Oligonucleotides for the cleavage assays

Upper strands

Name and length	Sequence
Substrat 1 (31mer)	(F)-5'-GGCTTATCTCCG**C<u>T</u>CGGG**TTAATCTGTCGCA
Substrat 2 (35mer)	(F)-5'-ACTTGGCTTATCTCCG**A<u>T</u>CGGT**TTAATCTGTCGCA
Substrat 3 (39mer)	(F)-5'-GGGTACTTGGCTTATCTCCG**CC<u>T</u>GGG**TTAATCTGTCGCA
Substrat 4 (43mer)	(F)-5'-GCTTGGGTACTTGGCTTATCTCCG**AC<u>T</u>GGT**TTAATCTGTCGCA
Substrat 5 (31mer)	(F)-5'-GGCTTATCTCCG**G<u>T</u>GCGC**TTAATCTGTCGCA
Substrat 6 (35mer)	(F)-5'-ACTTGGCTTATCTCCG**GCG<u>T</u>GC**TTAATCTGTCGCA
Substrat 8 (39mer)	(F)-5'-GGGTACTTGGCTTATCTCCG**A<u>T</u>CCCT**TTAATCTGTCGCA
Substrat 9 (43mer)	(F)-5'-GCTTGGGTACTTGGCTTATCTCCG**A<u>T</u>CAAT**TTAATCTGTCGCA
Substrat 10 (31mer)	(F)-5'-GGCTTATCTCCG**TC<u>T</u>CTA**TTAATCTGTCGCA
Substrat 11 (39mer)	(F)-5'-GGGTACTTGGCTTATCTCCG**TA<u>T</u>TAG**TTAATCTGTCGCA
FUT**A (31mer)	(F)-5'-GGCTTATCTCCGA**TT<u>T</u>AA**TTAATCTGTCGCA
FUAA (31mer)	(F)-5'-GGCTTATCTCCGA**GA<u>T</u>AC**TTAATCTGTCGCA
FUGxA (31mer)	(F)-5'-GGCTTATCTCCGA**GG<u>T</u>CA**TTAATCTGTCGCA
FUAG (35mer)	(F)-5'-ACTTGGCTTATCTCCGA**GA<u>T</u>GC**TTAATCTGTCGCA
FUCG (35mer)	(F)-5'-ACTTGGCTTATCTCCGA**GC<u>T</u>GC**TTAATCTGTCGCA
FUTG (35mer)	(F)-5'-ACTTGGCTTATCTCCGA**GT<u>T</u>CG**TTAATCTGTCGCA
FUAC (39mer)	(F)-5'-GGGTACTTGGCTTATCTCCGA**GA<u>T</u>CC**TTAATCTGTCGCA
FUGC (39mer)	(F)-5'-GGGTACTTGGCTTATCTCCGA**GG<u>T</u>CC**TTAATCTGTCGCA
FUT*T (43mer)	(F)-5'-GCTTGGGTACTTGGCTTATCTCCGA**TG<u>T</u>CT**TTAATCTGTCGCA
FUAT (43mer)	(F)-5'-GCTTGGGTACTTGGCTTATCTCCGA**GA<u>T</u>TC**TTAATCTGTCGCA

Lower strands

Name and length	Sequence
Substrat 1 (31mer)	5'-TGCGACAGATTAA**CCCGGG**CGGAGATAAGCC
Substrat 2 (31mer)	5'-TGCGACAGATTAA**ACCGGT**CGGAGATAAGCC
Substrat 3 (31mer)	5'-TGCGACAGATTAA**CCCGGG**CGGAGATAAGCC
Substrat 4 (31mer)	5'-TGCGACAGATTAA**ACCGGT**CGGAGATAAGCC
Substrat 5 (31mer)	5'-TGCGACAGATTAA**GCGCGC**CGGAGATAAGCC
Substrat 6 (31mer)	5'-TGCGACAGATTAA**GCGCGC**CGGAGATAAGCC
Substrat 8 (31mer)	5'-TGCGACAGATTAA**AGGGGT**CGGAGATAAGCC
Substrat 9 (31mer)	5'-TGCGACAGATTAA**ATTGGT**CGGAGATAAGCC
Substrat 10 (31mer)	5'-TGCGACAGATTAA**TAGGGA**CGGAGATAAGCC
Substrat 11 (31mer)	5'-TGCGACAGATTAA**CTAGTA**CGGAGATAAGCC
FLAA (31mer)	5'-TGCGACAGATTAA**GTGTC**TCGGAGATAAGCC
FLT**A (31mer)	5'-TGCGACAGATTAA**TTGAA**TCGGAGATAAGCC
FLxA (31mer)	5'-TGCGACAGATTAA**TGGCC**TCGGAGATAAGCC
FLAG (31mer)	5'-TGCGACAGATTAA**GCGTC**TCGGAGATAAGCC
FLCG (31mer)	5'-TGCGACAGATTAA**GCGGC**TCGGAGATAAGCC
FLTG (31mer)	5'-TGCGACAGATTAA**GCGAC**TCGGAGATAAGCC
FLAC (31mer)	5'-TGCGACAGATTAA**GGGTC**TCGGAGATAAGCC
FLGC (31mer)	5'-TGCGACAGATTAA**GGGCC**TCGGAGATAAGCC
FLT*T (31mer)	5'-TGCGACAGATTAA**AGGCA**TCGGAGATAAGCC
FLAT (31mer)	5'-TGCGACAGATTAA**GAGTC**TCGGAGATAAGCC

2.4.3.2 Oligonucleotides for the multiple substrate kinetics

Upper strands

Name and length	Sequence
Substrat 2 (35mer)	(F)-5'-ACTTGGCTTATCTCCG**ATCGGT**TTAATCTGTCGCA
FUT**A (31mer)	(F)-5'-GGCTTATCTCCGA**TTTAA**TTAATCTGTCGCA
FYT**A (31mer)	(F)-5'-GGCTTATCTCCGA**TTUAA**TTAATCTGTCGCA
FYT**A dark (31mer)	5'-GGCTTATCTCCGA**TTUAA**TTAATCTGTCGCA
MT2-9 (35mer)	(F)-5'-ACTTGGCTTATCTCCG**CCTGGG**TTAATCTGTCGCA
MU2-9 (35mer)	(F)-5'-ACTTGGCTTATCTCCG**CCUGGG**TTAATCTGTCGCA

MU2-9 dark (35mer)	5’-ACTTGGCTTATCTCCG**CCUGGG**TTAATCTGTCGCA
MT-3 (39mer)	(F)-5’-GGGTACTTGGCTTATCTCCG**CCTGGG**TTAATCTGTCGCA
MU-3 (39mer)	(F)-5’-GGGTACTTGGCTTATCTCCG**CCUGGG**TTAATCTGTCGCA
MU-3 dark (39mer)	5’-GGGTACTTGGCTTATCTCCG**CCUGGG**TTAATCTGTCGCA
FUGC (39mer)	(F)-5’-GGGTACTTGGCTTATCTCCGA**GGTCC**TTAATCTGTCGCA
FYGC (39mer)	(F)-5’-GGGTACTTGGCTTATCTCCGA**GGUCC**TTAATCTGTCGCA
FYGC dark (39mer)	5’-GGGTACTTGGCTTATCTCCGA**GGUCC**TTAATCTGTCGCA
FUT*T (43mer)	(F)-5’-GCTTGGGTACTTGGCTTATCTCCGA**TGTCT**TTAATCTGTCGCA
FYT*T (43mer)	(F)-5’-GCTTGGGTACTTGGCTTATCTCCGA**TGUCT**TTAATCTGTCGCA
FUT*T dark (43mer)	5’-GCTTGGGTACTTGGCTTATCTCCGA**TGTCT**TTAATCTGTCGCA

The sequences of the lower strands complementary to these oligos are presented above. For the preparation of the dsDNA substrates, following oligos were hybridised: “substrat 2” upper strand was hybridised to “substrat 2” lower; “FUT**A, FYT**A and FYT**A dark” – to FLT**A; “MT2-9, MU2-9, MU2-9 dark, MT-3, MU-3 and MU-3 dark” – to “substrat 3” lower; “FUGC, FYGC and FYGC dark” – to FLGC; “FUT*T, FYT*T and FYT*T dark” – with FLT*T.

2.4.3.3 Oligonucleotides to test the cleavage of different mismatches

For this, oligonucleotides MT-3, MT2-9, MU-3 and MU2-9 were used as upper, fluorescently labeled strands containing T or U bases, respectively. “MU-3 dark” and “MU2-9 dark” were also used as upper, non-labeled ones. “substrat 3 lower” was used as the lower strand with G opposition. Others are listed:

Upper strands

Name and length	Sequence
MUGC (31mer)	(F)-5’-GGCTTATCTCCGA**GGACC**TTAATCTGTCGCA
UpperA (31mer)	(F)-5’-GGCTTATCTCCG**CCAGGG**TTAATCTGTCGCA
UpperC (31mer)	(F)-5’-GGCTTATCTCCG**CCCGGG**TTAATCTGTCGCA
UpperG (31mer)	(F)-5’-GGCTTATCTCCG**CCGGGG**TTAATCTGTCGCA

Lower strands

Name and length	Sequence
LowerA (31mer)	5’-TGCGACAGATTAA**CCCAGG**CGGAGATAAGCC
LowerC (31mer)	5’-TGCGACAGATTAA**CCCCGG**CGGAGATAAGCC

LowerT (31mer)	5'-TGCGACAGATTAA**CCC<u>T</u>GG**CGGAGATAAGCC
LowerU (31mer)	5'-TGCGACAGATTAA**CCC<u>U</u>GG**CGGAGATAAGCC

For producing DNA duplex with MUGC upper oligonucleotide, FLGC lower strand (see above) was used.

2.5 Molecular weight markers

2.5.1 DNA size markers

The DNA size markers were mixed in 6x Loading Dye Solution (MBI, Fermentas) to the end concentration of 0.1µg/µl and stored at 4 °C.

GeneRuler^{TM} 1 kb DNA Ladder (MBI, Fermentas**)**: 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10.000 bp.

GeneRuler^{TM} 100 bp DNA Ladder (MBI, Fermentas**)**: 80, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 bp.

Lambda DNA/Eco47I Marker (MBI, Fermentas): 308, 310, 345, 398, 433, 511, 513, 590, 597, 894, 974, 985, 1284, 1420, 1611, 1951, 2005, 2134, 2555, 2606, 3676, 6442, 6555, 8126 bp.

2.5.2 Protein size markers

One of the markers used was the **Mid-range** marker (Promega). It contains following proteins with their respective weights:

Protein	Apparent molecular weight [Dalton]
Phosphorylase b	97,400
Bovine serum albumin	66,200
Glutamate dehydrogenase	55,000
Ovalbumin	42,000
Aldolase	40,000
Carbonic anhydrase	31,000

Soyabean Trypsin Inhibitor	21,500
Lysozyme	14,400

Another marker used was **10kDa Marker** (GibcoBRL): 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 200 kDa (exact composition was not provided by the producer).

2.6 Solutions and buffers

2.6.1 Solutions and buffers for the work with DNA

A.L.F. loading buffer

95% formamide, 20 mM EDTA, 3 mg/ml dextran blue

E1-buffer (GENOMED)
50 mM Tris/HCl pH 8.0, 10 mM EDTA, 100 μg/ml RNase.

E2-buffer (GENOMED)
0.2 M NaOH, 1 % (w/v) SDS.

E3-buffer (GENOMED)
3.2 M potassium acetate / acetic acid pH 5.5

E4-buffer (GENOMED)
600 mM NaCl, 100mM sodium acetate / acetic acid pH 5, 0.15% TritonX-100.

E5-buffer (GENOMED)
800 mM NaCl, 100 mM sodium acetate / acetic acid pH 5.

E6-buffer (GENOMED)
1250 mM NaCl, 100 mM Tris/HCl pH 8.5

Glycosylase reaction buffer, pH 7.5
20 mM Tris-HCl, pH 7.5, 20 mM $(NH_4)_2 SO_4$

Glycosylase reaction buffer, pH 9.0
20 mM Tris-HCl, pH 9.0, 20 mM $(NH_4)_2 SO_4$

6x Loading Dye Solution (MBI, Fermentas)
0.2% bromophenol-blue, 0.2% xylene cyanol, 60 mM EDTA, 60% glycerol.

P1 resuspension buffer (Qiagen)
50 mM Tris/HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNase.

P2 lysis buffer (Qiagen)
0.2 M NaOH, 1 % (w/v) SDS.

P3 neutralisation buffer (Qiagen)
3.0 M potassium acetate, pH 5.5.

QBT equilibration buffer (Qiagen)
750 mM NaCl, 50 mM MOPS, pH 7.0, 15% ethanol, 0.15% TritonX-100.

QC wash buffer (Qiagen)
1.0 M NaCl, 50 mM MOPS, pH 7.0, 15% ethanol.

QF elution buffer (Qiagen)
1.25 M NaCl, 50 mM Tris/HCl pH 8.5, 15% ethanol.

10x PfuTurbo DNA polymerase reaction buffer (Stratagene)
200 mM Tris/HCl pH 8.8, 20 mM $MgCl_2$, 100 mM KCl, 100 mM $(NH_4)_2 SO_4$, 1 % Triton^R X-100, 1 mg/ml BSA.

Restriction buffer R (MBI, Fermentas)
10 mM Tris/HCl pH 8.5, 10 mM $MgCl_2$, 100 mM KCl.

Restriction buffer G (MBI, Fermentas)
10 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl

Restriction buffer Y (MBI, Fermentas)
33 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 66 mM K-acetate.

Restriction buffer O (MBI, Fermentas)
50 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl.

Restriction buffer B (MBI, Fermentas)
50 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$.

SSC buffer
15 mM sodium citrate pH 7.2, 150 mM NaCl.

T4 ligase buffer (MBI, Fermentas)
40 mM Tris/HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP.

TBE-buffer
89 mM Tris, 89 mM boric acid, 2.5 mM EDTA

TE-buffer
10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0

10x Tlf DNA polymerase reaction buffer (Promega)
200 mM Tris-acetate (pH 9.0), 100 mM $(NH_4)_2 SO_4$, 750 mM potassium acetate, 0.5% Tween 20.

2.6.2 Solutions and buffers for the work with proteins

Alkaline phosphatase buffer
100 mM NaCl, 5 mM $MgCl_2$, 100 mM Tris/HCl, pH 9.5.

Anode buffer I
300 mM Tris/HCl, pH 10.4.

Anode buffer II
25 mM Tris/HCl, pH 10.4.

BB buffer
10 g dry milk, add PBS 1x till 100 ml.

BBT buffer

0.05% Tween 20 dissolved in 100 ml of BB buffer.

BCIP (5-bromo-4-chloro 3-indolyl phosphate)
0.5 g BCIP dissolved in the 100% dimethylformamide.

Cathode buffer
25 mM Tris/HCl, pH 9.4, 40 mM L-norleucin.

De-staining solution for PAGE
500 ml methanol, 100 ml acetic acid, add H_2O till 1l.

IMAC buffer
25 mM HEPES/KOH pH 7.6, 0.5 M NaCl, 0-1 M imidazole.

NBT (nitro blue tetrazolium)
0.5 g NBT dissolved in 10 ml of 70% dimethylformamide.

PBS buffer 10x
80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4, 2.4 g KH_2PO_4, pH 7.4, add H_2O till 1l.

PBST buffer
100 ml PBS 10x buffer, 0.05% Tween 20, add H_2O till 1l.

PBS/BBT buffer
25 ml PBS buffer 1x, 25 ml BBT buffer.

Protein sample buffer
62.5 mM Tris/HCl pH 6.8, 20 % (v/v) glycerol, 140 mM SDS, 3 % DTT, 0.1 % (w/v) bromphenol-blue.

Resolving gel buffer
1.5 M Tris/HCl pH 8.8

10x running buffer for PAGE
25 mM Tris, pH 8.3, 250 mM glycine, 0.5 % (w/v) SDS.

Stacking gel buffer
0.5 M Tris/HCl pH 6.8

Staining solution for PAGE
7.5% Coomasie Brilliant Blue R, 500 ml ethanol, 100 ml acetic acid, add H_2O till 1l

Working HEPES buffer
25 mM HEPES/KOH pH 7.6.

2.7 Enzymes and Proteins

Epicentre Technologies, Madison: *HK*™-UNG Thermolabile Uracil N-Glycosylase.

NEB New England Biolabs, Beverly: various restriction endonucleases**.**

MBI Fermentas, Vilnius: various restriction endonucleases, T4 DNA-ligase.

Promega, Madison: *Tlf* DNA Polymerase

Stratagene, California: *Pfu*Turbo DNA Polymerase.

Trevigen, Maryland: MutY glycosylase.

2.8 Chemicals

Applichem, Darmstadt: 30% Acrylamide, 2% Bisacrylamide solutions for A.L.F.-gel. Isopropyl-β-D-thiogalactopyranosid (IPTG).

Boehringer, Mannheim: Ampicillin, 2', 3'- Didesoxyribonucleoside-5'-triphosphate.

Qiagen GmbH, Hilden: Plasmid Midi-kit, penta-His tag antibodies.

Fluka Feinchemikalien GmbH, Neu-Ulm: Bromophenol-blue.

Genomed, Research Triangle Park: Jet-Star Plasmid kit.

Gibco/BRL GmbH, Eggenstein: Agarose electrophoresis grade, Yeast extract.

Macherey-Nagel GmbH & Co, Düren: Nucleotrap® kit.

Merck AG, Darmstadt: All other chemicals.

National Diagnostics, New Jersey: Protogel™.

Oxoid, Hampshire: Agar bacteriological grade, Trypton.

Pharmacia/LKB GmbH, Uppsala: Ammonium peroxodisulphate, N,N,N',N'-Tetramethylethylendiamine (TEMED).

Riedel de Haen AG, Seelze: Ethanol, Sodium chloride.

Serva Feinchemikalien GmbH & Co., Heildelberg: Ethylendiamine tetraacetic acid disodium salt (EDTA), Dithiotreitol (DTT), Kanamycin, Sodium dodecylsulphate (SDS), Urea.

Sigma Chemie GmbH, Deisenhofen: Coomasie Brilliant Blue R, Ethidium bromide, Dimethyl sulfoxide (DMSO), Anti-mouse antibodies IgG coupled with the alkaline peroxidase.

2.9 Instruments and Materials

Amicon®, Beverly: Centricon® and Centriprep® centrifugal devices.

Bender & Hobein AG: Vortex Genie 2.

Bio Rad, California: Gene Pulser II, E.coli Pulser™ Ciuvette.

Brand GmbH +Co, Wertheim/Main: Glass Pasteur pipettes, Cylinders.

Cybertech GmbH, Berlin: Cybertech CS-1 electrical ready picture camera.

Eppendorf-Netheler-Hinz GmbH, Hamburg: Eppendorf reaction tubes, Concentrator 5301.

Greiner & Söhne GmbH & Co KG, Nürtingen: Petri dishes, 10 ml and 50 ml glass tubes.

Heinemann Laboratoriums-Ausrüstungen, Schwäb. Gmünd: Branson Sonifier W-250.

Heinzinger, Germany: LNGs 350-06 power source.

Hettich Zentrifugen, Tuttlingen: Micro-litre bench centrifuge, cold rapid/k Microcentrifuge, Cold centrifuge Rotana/RPC, cold centrifuge Roto Silenta/RP.

INFORS AG, Bottmingen: Thermoshaker.

Kimbeley-Clark GmbH: Particles-free Kimwipes™.

Kontron Instruments GmbH, Eching: Two- ray spectrophotometer Uvikon 930.

MWG-Biotech, München: Thermal cycler.

PE Applied Biosystems, California: ABI PRISM™, 377 DNA Sequencer.

Perseptive Biosystems, Framingham: Vision™ Workstation, BioCad™, POROS columns.

Pharmacia/LKB , Uppsala: Activated Laser Fluorescence DNA-sequencer (A.L.F), Ultrospeck II, Blotting plates.

Sartorius, Göttingen: Minisart single use filter unit, Weight.

Savant Instruments, Bad Nauheim: Centrifuge RC-5C, Rotortype SS34

Schleicher & Schüll GmbH, Dassel: 3 MM- filter paper (Whatman), Disposable vacuum filtration system.

Terumo, Tokyo: syringes.

2.10 Software

Amplify ™ (Version 1.2)

Boxshade (Version 3.21)

www.ch.embnet.org/software/BOX_form.html

GCG Sequence Analysis Software Package (Version 3.2)

http://gcg.gwdg.de/cgi-bin/w2h/w2h.start

ClustalW

http://www2.ebi.ac.uk/clustalw/

CorelDraw (Version 7.0, Corel)

DNA Strider ™ (Version 1.2)

Fragment Manger VI.2™

SigmaPlot® (Version 5.00, SPSS Inc.)

TreeView (Version 1.5)

2.11 Databases

ExPasy Molecular Biology Server

http://www.expasy.ch/

National Center for Biotechnology Information

http://www.ncbi.nlm.nih.gov/

REBASE ™

http://rebase.neb.com

WIT2

http://integratedgenomics.com

3. Methods

3.1 Microbiological methods

3.1.1 Growth and storage of *Escherichia coli*

Bacteria were grown on the LB agar plates and in the liquid LB or 2 xYT media at 30-37 °C. All media and preparative tools were sterilised by autoclaving (20 min at 121 °C) with the exception of the heat labile compounds, which were filtered through a 0.2 µm filter (Sartorius). For the selective conditions appropriate antibiotics were added. For long-time storage, 0.9 ml of overnight bacteria cultures were mixed with 0.1 ml of 96 % sterile glycerol and placed into -70 °C.

3.1.2 Growth and storage of *Methanobacterium thermoautotrophicum*

Methanobacterium thermoautotophicum strain ΔH (DSM 1053) was obtained from the German strain collection (DSMZ Braunschweig). The culture was divided into four Hungate bottles with 100 ml of the special medium (Nölling *et.al*., 1991) in each. The medium was supplemented with the solutions of vitamins (Wolin *et.al*., 1964) and trace elements (Pfenning and Lippert, 1966). The cells were cultivated at 65 °C for three days with a twice-daily exchange of the H_2/CO_2 atmosphere.

For the following amplification procedures, one of the four bottles was opened and cells collected by centrifugation and re-suspended into 2 ml of TE buffer. The suspension was stored at -20^{o}C until needed.

The rest of the grown culture was stored in the anaerobic conditions, at 4^{o}C, in the dark.

3.2 Methods of molecular biology

3.2.1 Preparation of chemically competent cells of *Escherichia coli*

5 ml of 2 xYT medium was inoculated with a glycerol culture of the particular strain, shaken overnight at 37 °C, diluted 1:50 in a fresh 2 xYT medium and grown at 37 °C to an OD_{600} of 0.6. After centrifugation at 4000 rpm, 4 °C, 20 min, the cells were re-suspended in 20 ml 100 mM ice-cold $CaCl_2$ and incubated 20-24 h on ice. The

cells were collected by centrifugation; the pellet was re-suspended in 3 ml of 100 mM ice-cold $CaCl_2$ and incubated for 30 min on ice. This was divided into aliquots of 200 µl, 7 µl DMSO was added, mixed well and stored at -70 °C.

3.2.2 Preparation of the electrocompetent cells of *Escherichia coli*

50 ml of 2 xYT medium was inoculated with a glycerol culture of desired strain, shaken overnight at 37 °C, diluted 1:100 in 1l of fresh medium and grown at 37 °C to an OD_{600} of 0.6. The cells were collected by centrifugation at 6000 rpm, 4 °C 10 min. The pellet was re-suspended in the 100 ml of cold H_2O, centrifuged and again washed with 25 ml of cold H_2O. Finally, the cells were re-suspended in the 4 ml sterile 10% glycerol, divided into 50 µl aliquots and stored at -70 °C.

3.2.3 Transformation of the chemically competent cells of *E. coli*

For the transformation, 10-300 ng DNA was added to a 200 µl aliquot of the cell suspension. The cells were incubated for 30 min on ice, followed by 3 min heat shock at 37 °C. Then, they were mixed with 0.8 ml 2 xYT medium and incubated 40-60 min at 37 °C. 50-100 µl was plated out on the selective LB agar plates.

3.2.4 Transformation of the electrocompetent *E. coli* cells

1 µl of the fresh ligation reaction was mixed with 50 µl of competent cells, which were thawed slowly. After the transformation, the content of the tube was transferred into 1 ml 2xYT or SOC media and incubated at 37°C for 40-60 min. 250 µl was plated out on the selective LB agar plates.

3.2.5 Plasmid DNA preparation

Plasmid DNA prepared with the JetStar columns was used for the cloning and sequencing experiments. Plasmids purified by the Qiagen columns were used for the determining of the DNA methylation activity.

3.2.5.1 Plasmid Midi-preparation according Qiagen

50 ml of LB medium was inoculated with the desired *E. coli* and incubated at 37°C overnight with shaking. This was followed by centrifugation (6000 rpm, 4 °C, 15 min) and the pellet dissolved in 4 ml P1-buffer. Lysis of the cells followed immediate upon addition of 4 ml P2-buffer. After 5 min incubation at room temperature, 4 ml of P3 buffer was added to the lysate and then centrifuged (15000 rpm, 4 °C, 30 min). The supernatant was applied to a Qiagen-pack 100-column (Midi) that had been equilibrated before with 4 ml QBT-buffer. Column was washed twice with 10 ml QC-buffer and DNA eluted with 5 ml QF-buffer. Precipitation of DNA from the solution followed the addition of 0.7 volumes isopropanol and 30 min centrifugation at 15000 rpm, 4 °C. The DNA containing pellet was washed with 70 % ethanol and re-suspended in 50 ml TE-buffer.

3.2.5.2 Plasmid Midi-preparation according Jet-Star

50 ml 2 xYT medium containing the plasmid encoding *E. coli* strain was incubated at 37 °C overnight with shaking. This was followed by centrifugation at 6000 rpm, 4 °C, 15 min and the pellet dissolved in 4 ml E1-buffer. Lyses of the cells was started by the addition of 4 ml E2-buffer. Neutralization occurred upon the addition of 4 ml E3-buffer. Chromosomal DNA, proteins and membrane particles were removed after centrifugation 15000 rpm, 20 °C, 10 min. The supernatant was applied to a Midi JetStar-column, pre-equilibrated with 10 ml E4-buffer. The column was washed twice with 10 ml E5-buffer and DNA eluted with 5 ml E6-buffer. Plasmid DNA was precipitated with 0.7 volumes of isopropanol with the following wash of the pellet with 70 % ethanol. The DNA was re-suspended in 50 µl TE-buffer.

3.2.6 DNA agarose gel electropheresis

Analysis of DNA molecules was performed on the 0.8 - 1.5% agarose gel. Usually 10 µl of the product was mixed with 2 µl of 6x loading dye solution (MBI, Fermentas, Vilnius) and loaded on the gel. The separation of the fragments was done at constant voltage of 80V for 30-45 min in 0.5x TBE as a running buffer. Then the

gel was transferred to the bath with 0.5 mg/ml etidium bromide. After 10-20 min of staining the fragments were visualized at 302 nm UV light.

3.2.7 DNA purification from agarose gel by NUCLEOTRAP extraction kit

The DNA fragment is excised from agarose gel with a clean scalpel and transferred into the sterile tube. The DNA extraction from agarose gel was done by NUCLEOTRAP extraction kit essentially as described by producers. Solubilisation of agarose and binding of the DNA to the NUCLEOTRAP suspension was performed in three folds excess of N1 buffer at 50 ^{o}C. Non-bound components were removed by centrifugation. The remaining impurities were discarded by repeated washing and centrifugation steps with 500 µl N2 and N3 buffers, separately. The DNA was eluted from NUCLEOTRAP suspension in 50 µl of low salt buffer or water, pH 8.0-8.5 at 50-55 ^{o}C.

3.2.8 Determination of DNA concentration by UV-Spectroscopy

Nucleic acid concentrations were determined by the absorption of UV light at 260 nm. 1 OD_{260} corresponds to 50µg dsDNA and 33 µg ssDNA. The concentration of the oligos was calculated according the formula:

$$\mathbf{C\ (\mu M) = OD_{260}/(0.01 \times N)},$$

where N – number of bases

The purity of the sample can be estimated from the ratio of OD_{260} to OD_{280} (proteins absorb UV light at 280). A relative coefficient of about 2 (OD_{260}/ OD_{280}) is obtained from a protein-free DNA preparation.

3.2.9 Digest of DNA with the restriction endonucleases

Digestion of DNA was performed in the suitable buffer and under conditions described by the manufacturer. In general, about 0.1 µg DNA was digested with 2-5

units of a restriction endonuclease. When the digestion was done with two restriction endonucleases simultaneously, the best overall buffer was chosen. For the plasmid linerization 1-1.5 h of incubation were sufficient, for the removal the ends of PCR products - 12-16 h were needed.

3.2.10 Ligation of the DNA ends

Ligation of the sticky DNA ends was performed by the T4 ligase in the course of cloning into the pET vector. PCR products with the blunt end or with T/A overhangs were ligated into the TOPO vectors by action of the topoisomerase I.

3.2.10.1 Ligation by the T4 ligase

Typically, a ligation mixture included 20-50 ng vector DNA, insert in approximately 3-5 molar excess, 1 x T4 ligase buffer and 2–4 units of the T4 DNA ligase in a final volume of 20 µl. The samples were incubated overnight at 15 °C. A control reaction consisting of the above mentioned substances, except the insert was performed to estimate the re-ligation efficiency of the vector.

3.2.10.2 Ligation by the topoisomeraseI

In general, 0.5 µl to 4 µl of the typical PCR sample (10-20 ng/µl) was mixed with 1 µl of the linerized TOPO vector. Sterile water added till 5 µl final volume. The reaction was mixed gently and incubated 5 min at room temperature. After that, 1 µl of 6x TOPO Stop Solution was added (0.3 M NaCl, 0.06 M $MgCl_2$) and mixed well. 2 µl of that was transformed into the chemically competent TOP10 One Shot^{TM} cells. If the PCR product contains blunt ends, the pCR-Blunt II-TOPO vector was used. If the PCR was performed with the DNA polymerase without proofreading activity and has A overhang, pCR 2.1-TOPO was taken.

3.2.11 Purification and analysis of oligonucleotides by polyacrylamide gel electrophoresis (PAGE)

The 200 µl of samples were mixed with 1/3 (v/v) formamide and loaded on a 15% denaturing polyacrylamide gel. The separation was performed at 20W constant voltage for 4 hours, with 1 x TBE a running buffer. The desired oligonucleotide bands

were cut out of the gels and eluted by incubation at 42 °C in 500 µl TE buffer for 12 hr. The DNA was precipitated by the overnight incubation of the supernatant with 3 vol. of 96 % ice-cold ethanol. The DNA was re-suspended in 100 µl TE buffer or water.

3.2.12 Hybridization of oligonucleotides

Duplex DNA containing single base/base mismatches were prepared by hybridizing 6 pmol of a fluorescein labeled oligonucleotide to 30 pmol of a complementary, unlabelled oligonucleotide in 600 µl 1 x SSC buffer for 3 min at 80 °C and cooled down to the room temperature.

3.2.13 Polyacrylamide gel for the A.L.F run

For the preparation of the gel, the 60 ml of the following solution was made:

Reagent	Amount	End concentration
Urea	25.2 g	7 M
30% Acrylamide	21.3 ml	10 %
2% Bisacrylamide	10.7 ml	
10 x TBE	9.3 ml	1.5 x TBE
10 % APS	300 µl	
TEMED	50 µl	

Table 4. Composition of the polyacrylamide gel for the A.L.F run.

After all components, except TEMED and APS were dissolved, the solution was filtrated and degassed for about 10-15 min. Then, polymerisation was started by the addition of APS and TEMED and gel was poured between the prepared glass plates. Usual time of polymerisation was calculated to be between 1 and 1.5 hours.

3.2.14 Polymerase Chain Reaction (PCR)

PCR is a the repetitive bidirectional and exponential DNA synthesis via primer extension of a region of nucleic acid (Mullis and Faloona, 1987). Amplification of

certain DNA template requires two oligonucleotide primers, the four deoxynucleotide tripohosphates (dNTPs) and DNA polymerase to perform the synthesis. There are a lot of variations of PCR method described, depending on the type of DNA polymerase used, template as a starting material and desired product to be obtained (Dieffenbach and Dveksler, 1995).

During this work composition of PCR reaction was as following:

Reagent	End-concentration
Template DNA	1-50 ng
dNTP's	200 pmol each
Primers	0.5 pmol each
Reaction buffer	1 x
DNA polymerase	0.1 - 5 U
Mg^{+2} (chloride or sulfate)	1.5 mM
DMSO when needed	5%

Table 5. Standard composition of the PCR reaction.

The amount of analytical reaction was set up to be 20 μl, the preparative - 50-100 μl.

After the initial denaturation step at 94-96 ^{o}C for 1-2 min, the amplification was usually repeated 25-30 cycles and later cooled down to 4 °C.

Segment	Time	Temperature
Denaturation	30-45 sec	92-94°C
Annealing	30 sec -1 min	45-68°C
Extension	1-2 min	72°C

Table 6. Cycling profile for a typical amplification reaction.

The temperature and prolongation of the denaturation step depends on the DNA polymerase used: *Tlf* and *Tag* polymerases being less, *Pfu* and *Vent* polymerase - more thermostable. The annealing step depends mainly on the GC-content of the template and was kept 2-5 degrees lower than Tm of the primers.

Elongation step was programmed due to the content and length of the DNA template. (Dieffenbach and Dveksler, 1995).

3.2.14.1 Screening of the colonies by PCR

Colonies were inoculated into 3 ml of 2xYT media with the appropriate antibiotic and grown at 37°C overnight or over day. 10 µl of the culture was transferred into the tube, 90 µl sterile water was added, mixed vigorously and incubated at 95°C for 15 min. The mix was shortly centrifuged and placed on ice. 1 µl of it was used for the following PCR reaction. Typically, the amplification was performed in a 20 µl volume with *Tlf* polymerase, suitable buffer with $MgSO_4$ and the primer pair, specific for the insert or vector.

3.2.15 DNA sequencing

2.5 µl of purified plasmid DNA (250-400 ng of DNA) was mixed with 0.5 µl of appropriate primer (end concentration 5 pmol), 0.5 µl DMSO, 4 µl of BigDye™ premix (containing modified Taq DNA polymerase, buffer, nucleotides and dye-terminators) and adjusted till 10 µl with sterile water. The sequencing reaction was performed according similar conditions, described in the Table 6. For majority of the genes, sequenced in this work, annealing temperature of 52 °C, extension – 60 °C was used. To precipitate residual dye terminators, 75 µl of freshly prepared ethanol/sodium acetate solution (consisting of 9 ml 96% ethanol, 1.8 ml water, 0.3 ml 3M sodium acetate, pH 4.6) was added. The mix was incubated 15 min at room temperature, centrifuged 30 min at 3300g and decanted. Then, the pellet was dissolved in 9 µl of water and 1 µl of stop-mix (85% formamide, 25 mM EDTA, pH 8.0, 50 mg/ml Dextran blue). Following incubation at 60 °C for 1h was resulting in simultaneous denaturation and reduction of the sample volume. 0.5 µl of that was loaded on the sequencing gel. Alternatively, dye terminators were removed by Sephadex G-50 Microspin columns (Pharmacia Amersham).

The separation of the DNA fragments was performed on the vertical ABI Prism™ DNA sequencer 377. The gel consisted of following components: 18g urea, 5.3 ml 40% PAGE-Plus (Amresco), 6 ml 10x TBE (tris-boric) or TTE (tris-taurine) buffer, 21.5 ml water. The components were well mixed, filtrated and degassed for 10-15 min. Polymerization was started with by addition of 250 µl 10% APS and 25 µl TEMED. The gel was polymerized for 1h at room temperature. The run was

performed in the 2x running buffer in the upper buffer reservoir, 1x - in the lower, under following conditions: currency 21.0 mA, voltage 46 W, 40 mW laser power, 45 ^{O}C temperature. Sequence Analysis Software (version 3.2) was used to track the gel. LR-377 base caller was implemented to read the sequencing data.

3.3 Methods to work with proteins

3.3.1 SDS-Polyacrylamide Gel Electrophoresis (PAGE-SDS) of Proteins

Proteins were separated on a discontinuous polyacrylamide gel system as described by Laemmli, 1970. Inclusion of SDS denaturant allowed of the proteins depending only on their size (Scope, 1993).

Gels used for this work, usually contained 15-20 % (v/v) of acrylamide. The composition of the gels are shown in the following table:

Stacking gel	AA	0.5M Tris	SDS	APS	TEMED	H_2O
5 %	3.34 ml	2.5 ml	200 µl	200 µl	10 µl	14 ml
Resolving gel	AA	1.5M Tris	SDS	APS	TEMED	H_2O
15 %	40 ml	15 ml	800 µl	300 µl	40 µl	24.6 ml
20 %	53 ml	15 ml	800 µl	300 µl	40 µl	11.2 ml

Table 7. Solutions to prepare six 15-20% discontinuous SDS-polyacrylamide gels.

After casting the resolving gel, isopropanol was added on to the top to give a level surface. After polymerisation of this gel, the isopropanol was decanted and the stacking gel was applied in to the chamber.

1/2 volume of protein sample buffer was added to the samples and followed by 10 min incubation at 95 °C for denaturing the proteins. Prepared samples were loaded on the gel and electrophoresis was performed at a constant current of 25 – 40 mA. After electrophoresis, the protein bands were visualised by emerging the gel into the staining solution for 20-30 min. Alternatively, after the electrophorese proteins can be transferred from the gels onto the nitrocellulose filter for the immunochemical detection.

3.3.2 Determination of protein concentration

The protein concentration was measured by absorbency at 280 nm (A_{280}) and calculated according to Lambert-Beer's law:

$$\mathbf{OD_{280} = \varepsilon_{280} \times d \times c}$$

where: ε_{280}: molar extinction coefficient ($M^{-1}cm^{-1}$)
d: path-length determined by cuvette thickness (cm)
c: protein concentration (M)

The absorbance of a protein solution above 275 nm depends only on the several chromophores: tryptophane (*Trp*), tyrosine (*Tyr*) and cysteine (*Cys*) (Wetlaufer, 1962). According this, Pace *et.al*. 1995, derived the formula for the calculation of the extinction coefficients of the globular proteins, dissolved in a water and containing *Trp* residues:

$$\mathbf{\varepsilon_{280} = 5500 \times (no\ of\ Trp) + 1490 \times (no\ of\ Tyr) + 125 \times (no\ of\ Cys)}$$

As all the proteins described in this work satisfied the limitations, consequently, this formula was used for the calculation of the $\varepsilon_{280.}$

3.3.3 Purification of proteins

The proteins investigated in this work were purified essentially as described below. If there were some differences from the main protocol, they are detailed in the "Results and Discussion" section.

E. coli BL21(DE3)pLysS (Novagen) harbouring pET-21d plasmid with respective insert was grown overnight at 37 °C in 50 ml 2xYT medium containing 50 μg/ml ampicillin and 34 μg/ml chloramphenicol. 10 ml overnight culture were transferred to 1 l fresh medium containing antibiotics and grown at 37 °C until the OD_{600} reached 0.6. IPTG was added to a final concentration of 1 mM and the culture was incubated for additional 3 h at 30 °C. Cells were harvested by centrifugation, re-suspended in 20 ml 0.5 M NaCl, 25 mM HEPES-KOH, pH 7.6 and frozen at -70 °C. Cell lyses was induced by rapidly thawing in a 25-37 °C water bath and sonication on ice (Branson sonifier, 10 bursts of 10 sec each). The crude lysate was clarified by centrifugation at 15,000xg at 4 °C for 20 min and applied to a column filled with 3 ml

of Chelating Sepharose^{TM} Fast Flow (Amersham Pharmacia Biotech) loaded with Ni^{2+} (IMAC). All proteins described in this work were eluted with 200 - 300 mM imidazole concentration in 0.5 M NaCl, 25 mM HEPES-KOH, pH 7.6.

The elute was diluted ten-fold with cold 25 mM HEPES-KOH, pH 7.6 and loaded onto a heparin (Mig.*Mth*II, MutY.*Tth*) or HS columns (TTUDGA, X216) (POROS 20 columns, 4.6 mm x 100 mm). The proteins were eluted in a linear gradient of 0-1.5 M NaCl (BioCad^{TM} Workstation, PerSeptive Biosystems) and analysed on SDS-PAGE to be homogeneous. Appropriate fractions were pooled, desalted and concentrated 10 times by ultrafiltration (Millipore Centriprep cartridge, molecular weight cut-off ca. 3000). The resulting solution was diluted with an equal amount of glycerol and DTT was added to a final concentration of 1 mM. The enzymes were stored -20 °C.

3.3.4 "Western blot" with the anti-His tag antibodies

The experiment was done mainly according the Sambrook *et.al*., 1989. The changes from the standard protocol were introduced according the comments of Dr. R. Dietrich (personal communication).

3.3.4.1 Transfer of the proteins on the nitroceliulose membrane

After the separation of the proteins on the PAGE-SDS (**3.3.1**), the form was disassembled and the gel was transferred in the 100 ml of the Cathode buffer and incubated for 5 min. Meanwhile the membrane was equilibrated with the Anode buffer I for 10 min. The three layers of the filter paper were emerged into Cathode buffer, two – into Anode buffer I, and one – into the Anode buffer II. The filter soaked in the Anode buffer I was placed on the anode plate. On the top of that, filter with the Anode buffer I was placed, then nitrocellulose membrane and the protein gel. Finally, filter with the Cathode buffer was laid and covered with the cathode plate. The transfer was done for about 1 hour at 100 mA constant current.

3.3.4.2 Imunochemical detection of the proteins

After the transfer of the proteins onto the membrane, the blocking of the binding sites was performed overnight in the 60 ml of the Blocking buffer (BB) at 4 ^{o}C temp. Binding of the monoclonal penta-His tag antibodies (Qiagen) was done by incubation in the 20 ml BBT buffer for 2h at the room temperature. Final

concentration of the antibodies was 0.1 μg/ml. Then, the membrane was transferred into the 100 ml of PBST buffer and incubated for 10 min at room temp. The wash was performed in the PBS buffer 4 x 5 min and the membrane was dried for 20 min. On the dried membrane 40 ml of the PBS/BBT buffer was applied with the anti-mouse antibodies IgG, coupled to the alkaline phosphatase with the final concentration of 0.1 μg/ml. Incubation was done for 1 hour at room temp. The wash step with the PBST and PBS buffers was repeated and the membrane was dried again. The staining was performed with the BCIP/NBT substrate. The process was stopped by the addition of the PBS buffer with the 20 mM EDTA.

3.4 Enzymatic activity tests

3.4.1 Procedures for the glycosylase and AP-cleavage analyses

The standard reaction mixture consisted of 40 fmol labelled substrate (*i.e.* 4 μl of duplex stock solution prepared as described in the preceding paragraph) and the appropriate enzyme in the glycosylase reaction buffer. The exact conditions, like amount of the respective enzyme, pH of the reaction buffer and *etc*. are indicated in the appropriate “Results and Discussion“ sections. In general, the reactions were carried out at 50 °C for 5-60 min. AP-sites were cleaved by adding 2 μl of 1 M NaOH and keeping the mixture at 95 °C for 5-10 min (Horst and Fritz, 1996). For gel electrophoretic analysis, 10 μl of A.L.F. loading buffer were added and samples were loaded onto a denaturing polyacrylamide gel (Pharmacia A.L.F. DNA sequencer) in 10 μl aliquots. For assaying glycosylase-associated AP lyase activity, the alkali/heat treatment was omitted.

If needed, the prepared reactions could be stored at –20 °C, in the dark for several months.

3.4.2 Multiple substrate kinetics

The proceedings of multiple substrate kinetics are about the same as that of normal cleavage assays except that more than one substrate is present in the reaction mixture. Usually, the mix of 2-3 substrates with the varying length was used. The reactions were mainly done at 40-50 °C to reduce the speed of the reaction. 20 μl probes were taken out at the different time points and subjected to the alkali/heat treatment. Normally, 10-12 time points were taken for the calculation of the one

curve. The time interval was adjusted for the every enzyme individually and is indicated in the “Results and Discussion” section.

3.5 Methods of computer work

3.5.1 Sequence homology searches by BLAST

Search for similar sequences was performed by BLAST (Basic Local Alignment Search Tool) (Altschul *et.al*., 1997). Protein sequences of interest were compared to nucleotide databases, translated in six frames: tblastn option. Low complexity regions of query sequence were filtered. As default scoring matrix BLOSUM62 was used, default word size was chosen 3, gap opening costs 11, gap extension costs 1. Into consideration was taken the sequences producing hits with e-value smaller than 0.01. BLAST was performed in the databank of National Centre of Biotechnological Information (NCBI) and internal database of unfinished genomes of Göttingen Genomics Laboratory.

3.5.2 Calculation of sequence homology and similarity

Expression of identity and similarity between two protein sequences in percents was obtained by BLAST 2 sequences (Tatiana *et.al*., 1999) option at NCBI. Parameters of the pairwise alignment essentially were the same as described before. Alternatively, GCG Sequence Analysis Package was used with the option GAP. Then, penalty for gap opening was 8, extension – 2, while the matrix used was the same BLOSUM62.

3.5.3 Multiple sequence alignments and phylogenetic trees

Multiple sequence alignments were created by ClustalW program (Thompson *et.al*., 1994) at European Bioinformatics Institute EBI under default parameters. Alternatively, ClustalX (Thompson *et.al*., 1998) was used. The latter implied mode of Multiple Sequence Alignment, BLOSUM matrix, costs for gap opening 10, for gap extension – 0.05. Delay of divergence was 40%. In cases of both programs, poorly conserved N- or C- terminals of protein sequences were removed.

The multiple sequence alignments were used to calculate phylogenetic trees by Neighbour-Joining method and displayed by TreeView program (Page, 1996).

3.5.4 Work on sequencing data of *Thermus thermophilus* HB27 genome

Work on the appropriate sequence contigs was done at Staden Package according producers manual (Bonfield *et.al*., 1998). Editing and alignment of the raw sequences was performed in GAP4 subprogram. Search for open reading frames, nucleotide sequence translation into protein sequences and determination of start and stop codons were done in NIP4 subprogram.

4. Results and Discussion

4.1 Characterisation of Mig.*Mth*II from *Methanobacterium thermoautotrophicum* ΔH

The project started from homology search of U/G, T/G specific activities in the sequencing data of *Methanobacterium thermoautotrophicum* ΔH at NCBI database. As queries protein sequences of glycosylases, listed in Table 1 were used. Search revealed the presence only of glycosylases of helix-hairpin-helix family.

4.1.1 Identification of Mig.*Mth*II as a member of Mig glycosylase subfamily

To differentiate among the identified *orf*s in *M. thermoautotrophicum*, they were compared with the sequences of the other helix-hairpin-helix family glycosylases originating in thermophiles. For this, identified *orf*s were aligned by Clustal X, the unrooted phylogenetic tree was calculated and displayed by the TreeView program (Page, 1996) (Fig.1). Majority of appropriate genes were found at NCBI database of the annotated microbial genomes. The sequences of unfinished *Thermus thermophilus* HB27 were taken from the WIT2/restricted database. Sequence of PA.Mig was added after the recent publication (Yang *et.al*., 2000). To obtain well-separated subtypes of the family, only the closest homologs, MutY and Endo III-like sequences were included into the alignment.

As it is seen from Fig 1, the proteins cluster clearly into MutY or Endo III–like subtypes. U/G, T/G specific Mig.*Mth* is more similar to MutY-type sequences, and, importantly, is not the only representative any more. The highest degree of similarity was found to *orf* originally annotated as an endonuclease III (Smith *et.al*., 1997), which was submitted into the public databases under the name MTH 496. Latter was identified on the chromosome of archaeon *Methanobacterium thermoautotrophicum* ΔH. Later, the purified and experimentally tested protein was named Mig.*Mth*II due to its similarity to Mig.*Mth* (now called Mig.*Mth*I). Other *orf*s of *M. thermoautotrophicum* ΔH (MTH 746, MTH 764) stand well apart from Mig.*Mth*I (Fig. 1) and, very likely, code for different, yet unknown specificities.

The next homolog of Mig.*Mth*I (APE 0875) was found in the recently sequenced archaeal genome of *Aeropyrum pernix* (Kawarabayasi *et al*., 1999). Finally, Yang *et.al*., 2000 reported the U/G, T/G activity of the one more helix-hairpin-helix glycosylase, called PA.Mig.

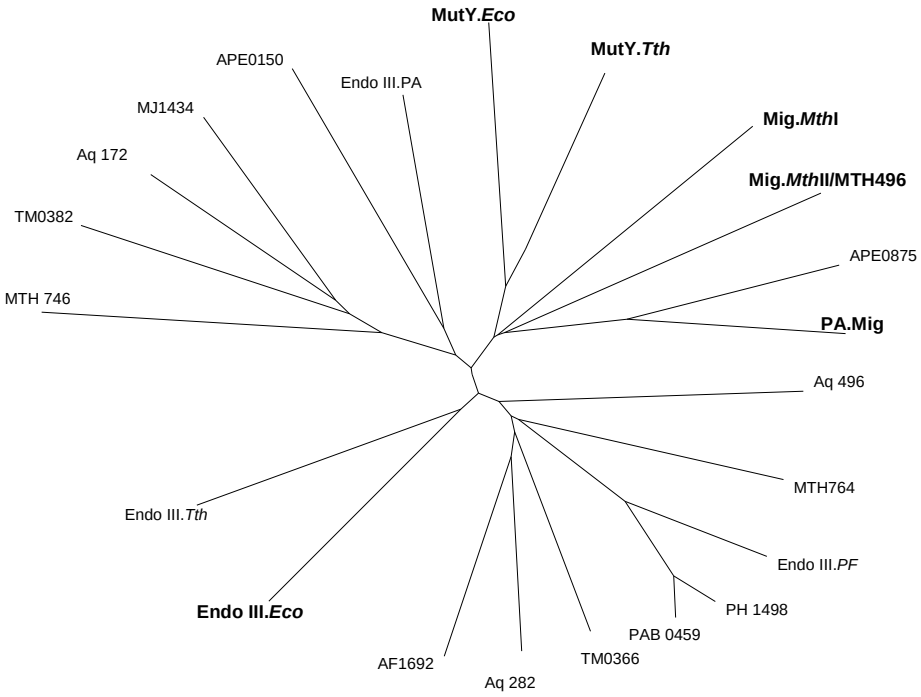


Fig 1. Phylogenetic tree of the thermostable homologs of the hairpin-helix-hairpin glycosylase family. Aq – *Aquifex aeolicus*, AF – *Achaeglobus fulgidus*, APE – *Aeropyrum pernix*, *Eco* – *Escherichia coli*, MJ – *Methanococcus jannaschii*, MTH – *Methanobacterium thermoautotrophicum*, PA – *Pyrobaculum aerophilum*, PAB – *Pyrococcus abyssi*, PF – *Pyrococcus furiosius*, PH – *Pyrococcus horikoshi*, TM – *Thermotoga maritima*, Tth – *Thermus thermophilus*. Marked names represent experimentally tested activities.

Despite the high similarity to the other members of the family (about 30% identity, 40-50% similarity), Mig enzymes share several specific features, which allow their easy clustering into the same subfamily (Fig. 1). First, all of them are found in close proximity to the 5-cytosine specific DNA methyltransferases. Second, they do not contain C-terminal extension of their sequence: this distinguishes them from MutY. As reported by Yang *et.al*., 2000, the enzymes contain conserved amino acid residues in the active pocket, specific only for the U/G, T/G glycosylases. All these features enabled separation of the enzymes into the helix-hairpin-helix subfamily, called Mig-type glycosylases.

Homology search of the Mig-glycosylases was performed in the WIT2/restricted database and additional members were found among unpublished data. The multiple sequence alignment was constructed to get more information about the features specific only for the Mig-glycosylases and the relatedness of the proteins within the subgroup (Fig 2). Due to similarity to the experimentally characterised Mig.*Mth*I and PA.Mig, it is reasonable to expect that the other *orf* also code for the U/G and T/G specific enzymes. Therefore, the names with the Mig specification were introduced.

The putative enzymes shear several highly conserved patterns characteristic to the helix-hairpin-helix family (Thayer *et.al*., 1995) (red colour in Fig. 2). The violet colour indicates the proposed regions for uracil and thymine recognition (Guan, *et.al*., 1998; Fondufe-Mittendorf, personal communication). The green colour possibly differentiates the Mig glycosylases due to their various sequence selectivity. More detailed discussion about the role of the specific residues in the substrate recognition and catalysis is presented in **4.1.10** and **4.2.12**.

Thus, *orf* MTH 496 is the only candidate for the repair of U/G, T/G mismatches in *M. thermoautotrophicum* ΔH to date. Therefore, it was chosen for cloning and characterisation of the protein, with particular emphasis on elucidation whether it could function as a general defence mechanism for the whole genome.

4.1.2 Re-annotation of *mig*-2 gene (*orf* MTH 496)

The multiple sequence alignment revealed that MTH 496 is the shortest protein among the family of hairpin – helix – hairpin glycosylases (Fig 2). The average length of most of the proteins is about 220 a.a. MTH 496, however, contains only 175 a.a, thus missing about one/fifth of the “normal” length.

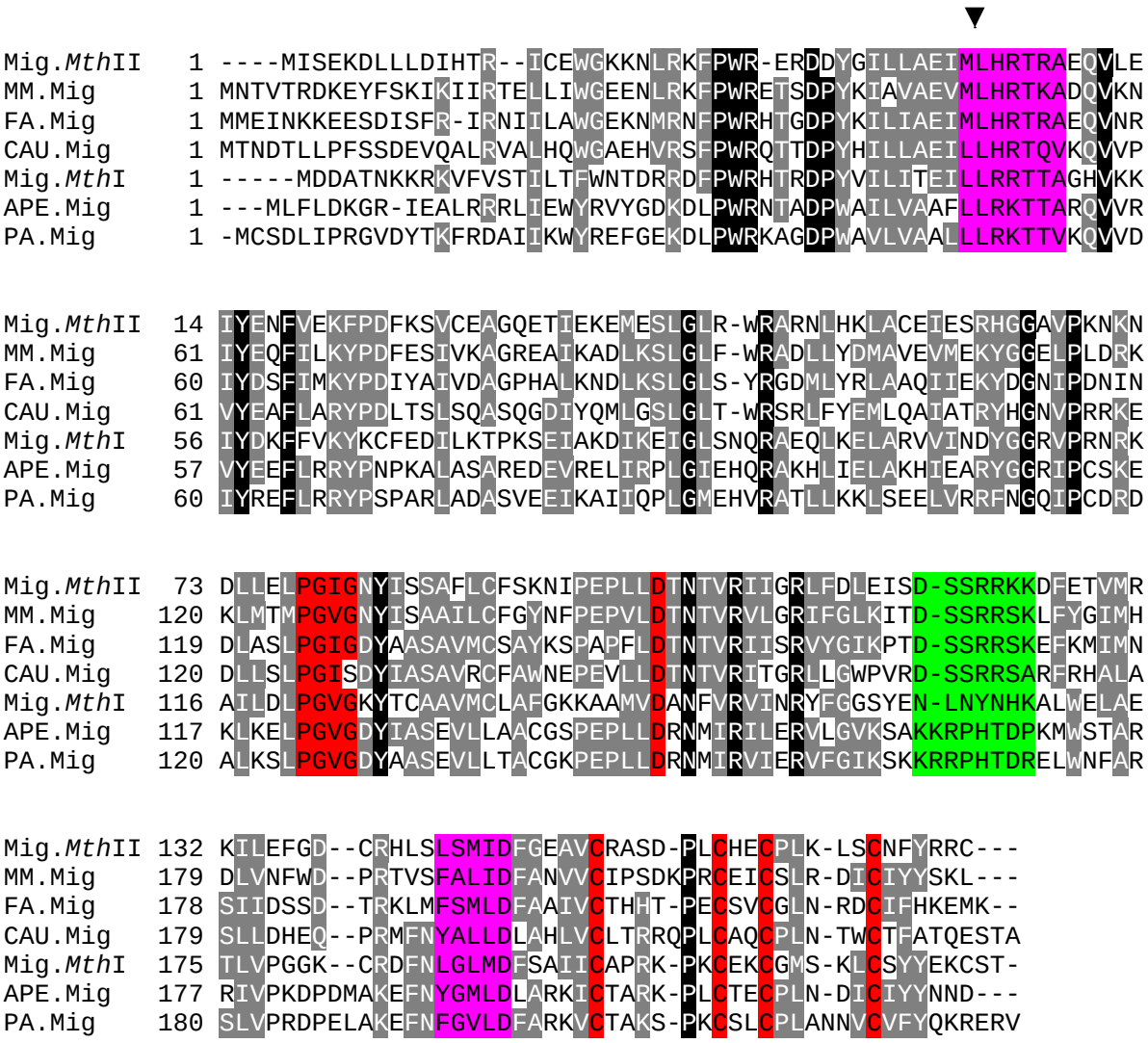


Fig 2. The multiple sequence alignment of the Mig glycosylases. APE – *Aeropyrum pernix*, CAU – *Chloroflexus aurianticus*, FA – *Ferroplasma acidophilum*, MM – *Methanosarcina mazei*, PA– *Pyrobaculum aerophilum*. Black colour indicates the residues, which are fully conserved within the family, grey – similar in all proteins or conserved in the four of the representatives. The red colour symbolises the features specific to the helix-hairpin-helix glycosylase family, violet – characteristic to the Mig subtype, green – possible variable region responsible for the sequence specificity. Black triangle indicates the start of Mig.*Mth*II (*orf* MTH 496) as originally annotated by Smith *et.al*., 1997 (for more see text).

Recent publication on the crystal structure of the catalytic core of the MutY from *Escherichia coli* (Guan *et al*., 1998) presented detailed description of the motifs required for the substrate recognition. One of them is " EVMLQQTQV", which is

located close to the N-terminal part of the enzyme. The respective region among the Mig glycosylases is "LLRKTTV" (violet colour in the Fig. 2). According to the original annotation (Smith *et.al*., 1997), MTH 496 starts from methionine "*Met*", which already is within the motif. Therefore this start is unlikely due to the overall sequence similarity of MTH 496 to the other Mig glycosylases and to the MutY. Thus, the upstream sequence of the MTH 496 was checked in the publicly deposited sequences of the *M. thermoautotrophicum* ΔH (accession number AE000833). The respective position of the contig contains several polyA and polyG tracks. Therefore, in order to avoid any possible mistake in the deposited sequences, the upstream region of the MTH 496 was re-sequenced. For this, the region was amplified by PCR from the cell suspension of the *M. thermoautotrophicum* (**4.1.4**). Expand^{TM} High Fidelity PCR System and the primer pair MTHconf /496Xho were used. The resulting PCR product was purified and sequenced on both strands (Fig 3).

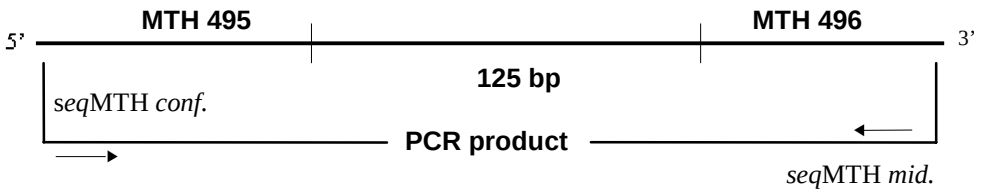


Fig 3. Scheme of re-sequencing the upstream region of the *orf* MTH 496. *seq*MTH*conf* and *seq*MTH*mid* – primers for the sequencing of the PCR product.

The sequence obtained is the perfect match with the one deposited in the database. Then it was translated into protein sequence by the **DNA Strider** program into six frames. Each putative protein was screened for the presence of the patterns, characteristic to *orf* MTH 496. The respective sequence was identified, extended upstream till the first stop codon located in frame and aligned to Mig.*Mth*I. The alignment showed unambiguously that there was a mistake in the original annotation of the *orf* 496 (black arrow in Fig. 2).

However, complications arose in defining the full-length product as no "*ATG*" ("*Met*" starting codon) was found "upstream" part of the sequence. To solve this, the statistical data of the usage of the start codons in the whole genome of the *Methanobacterium thermoautotrophicum* (Smith *et al*., 1997) were compared with the

distribution of the putative start codons in the re-annotated sequence of the MTH 496 (Table 8). As seen, the most frequent start codon is "*ATG*", but the use of "*GTG*" and "*TTG*" is also significantly high in *M. thermoautotrophicum.* This is, in a good agreement with the data published for the archaeal microorganisms (Dennis, 1997). However, as seen from Table 8, not only "*ATG"*, but also "*GTG"* and "*TTG"* were not found in the upstream part of mig2 gene. Thus, different codon acting as a start needed to be identified. As the most conceivable candidate codon "*CTG*" was suggested. It was chosen due to inverse wobbling, which would allow recognition of a start by proper pairing of the second and third nucleotides of the codon with the anticodon (UAC) of formylmethionine tRNA. First nucleotide U then, could make interactions not only with A, but also with G, T and, probably, C. Thus, "*CTG"* could possibly act as a start signal. As many as five "*CTG"* codons are located in this re-annotated region (AE000833). The most left side removed "*CTG*" was chosen as a putative start for the full-length protein (Fig. 2, black arrow).

Start codons	**Distribution in the whole genome of *M.thermoautotrophicum* (in percents)**	**Distribution in the upstream part of mig-2 gene (in numbers)**
ATG (Met)	63	0
GTG (Val)	22	0
TTG (Leu)	15	0
CTG (Leu)	0	5

Table 8. Distribution of start codons in the genome *M. thermoautotrophicum* ΔH.

Clearly, the assignment of the new start required more evidence. It has been described that archaeal organisms use the Shine-Dalgarne sequence to recognise the start codon in more that half of their genes (Dennis, 1997). Sequence of rbs (ribosome binding site) of methanogenic archaea strongly resembles the canonical procaryotic rbs: 5' AGGAGG 3', which is located in the 3-10 nucleotides distance from the starting codon (Brown *et al*., 1989). Accordingly, the search of the appropriate sequence pattern was performed in the "upstream" region of the MTH 496 by GCG Sequence Analysis Software Package. There was no positive output with the "AGGAGG" string as a query sequence. However, repeated iteration with more relaxed patterns resulted in hit 5 nucleotides upstream the new start codon "*CTG"* (Fig. 4). No signal was obtained in front of the originally assigned "*ATG"* codon.

Based on this, the chosen "*CTG"* codon was assigned as a start for the full-length Mig.*Mth*II and the gene of the respective length was cloned for the further research (**4.1.4**).

In general, "*CTG"* is expected to be of the most rarely used start codons, though some examples have been documented, like in equine infectious anemia virus (Carroll and Derse, 1993). After the re-annotation of *mig2* gene, the absence of "*CTG*" as a start in the *M. thermoautotrophicum* (Smith *et.al*., 1996) (Table 8) looks suspicious. Most likely, that the latter option was not included in the gene search program, thus the re-annotation of the whole genome of *M. thermoautotrophicum* would be highly advantageous.

4.1.3 Re-annotation of *mig*2 locus

Correction of the length of the MTH 496 revealed that the *orf* is not only located in the close proximity to DNA methyltransferase MTH 495, but the genes of both enzymes indeed overlap. The most probable case is that stop codon of the DNA methyltransferase MTH 495 "*TGA*" is part of the start codon of the glycosylase MTH 496 "*CTG*" (Fig 4).

A somewhat similar situation was described for the Vsr/Dcm in *Escherichia coli* (Sohail *et.al*., 1990). Apparently, Vsr and Dcm are transcribed as a single unit, but separate proteins are produced (Dar and Bhagwat, 1993). In general, much less is known about the transcription/translation regulation in archaea (Olsen and Woese, 1997). Therefore, the mode of the regulation of the MTH 495/MTH 496 gene couple is currently speculative. Nevertheless, the possible rbs "AGGGGG" can be detected in front of the glycosylase gene (Fig 4). The string "AGAGGG" is present close to the DNA methyltransferase gene (data not shown). It is, therefore, most likely, that the proteins are translated as separate units.

Another methanogenic archaeon *Methanosarcina mazei* also contains the gene of Mig glycosylase (MM.Mig) (Fig. 2), which also is located in close proximity to the putative C5-cytosine specific methyltransferase. Surprisingly, both genes have the same localization as in *M. thermoautotrophicum*, namely, stop codon of methyltransferase is a part of start codon for Mig glycosylase with the +1 frameshift (data not shown). In case of *M. mazei*, Mig encoding gene starts from "*ATG"* codon. In addition, analysis of the "upstream" region of the MM.Mig glycosylase of also revealed the presence of a putative rbs sequence in 4 nucleotides distance from the start codon ("AAGAGA"). Alignment of the nucleotide sequence between MM.Mig and Mig.*Mth*II also showed the high degree of conservation the 5' non-coding regions

of the respective genes (data not shown). Listed findings point at close evolutionary relatedness between two archaeal organisms and also lend further support to the assignment of the reading frame for Mig.*Mth*II.

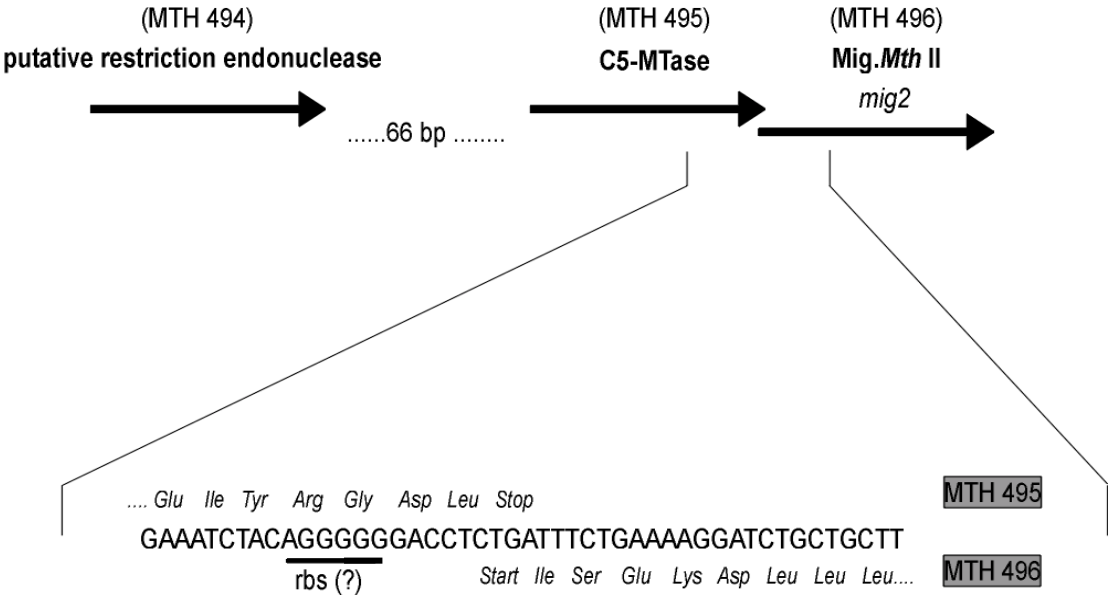


Fig 4. Organisation of the *mig*2 gene locus in the chromosome of *M. thermoautotrophicum* ΔH. Reading frames are numbered as assigned originally (Smith *et.al.*, 1997). "C5-Mtase stands for "DNA-cytosine 5-methyltransferase". Arrows indicate direction of transcription. A possible ribosome binding site (rbs) for translation of *mig2* mRNA is underlined. "CTG" was chosen as the most plausible start codon of *mig2* mRNA.

4.1.4 Cloning of the full-length and truncated Mig.*Mth*II

Methanobacterium thermoautotophicum strain ΔH (DSM 1053) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ Braunschweig) and cultivated in a special medium under anaerobic conditions as described in Methods. The growth and following confirmation of the identity of the microorganism under a fluorescence microscope was done with the help of Dr. Uwe

Deppenmeier and Dr. Hans-Jörg Abken from Institute Mikrobiology and Genetik, Georg-August Universität Göttingen.

As the preparation of genomic DNA of *M. thermoautotrophcum* is known to be technically difficult and inefficient procedure (Horst, 1996), the amplification was performed directly on the cell homogenate. The cell walls were disrupted by 15 cycles of heating at 95 °C for 10 min and quick freezing at -70 °C for 2-3 min. The treatment resulted in a crude, viscous homogenate that was used directly for PCR amplification of the appropriate genes.

The PCR was carried out with the oligonucleotide pair 496Nco and 496Xho for the truncated Mig.*Mth* II gene; Ncochrom and 496Xho - for the full-length gene. As it is seen from the names, primer 496Xho contains an *Xho*I, primers Ncochrom and 496Nco each an *Nco*I restriction site for the subsequent cloning into the expression vector pET21d (Novagen). The amplification was done with the Expand^{TM} High Fidelity PCR System (Boehringer Mannheim GmbH) with 60 ^{o}C as an annealing temperature. Products of desired length were obtained (540 b.p. for truncated, 663 b.p. for full-length *mig*-2).

DNA of the PCR products was purified with the NUCLEOTRAP-kit, digested with the appropriate restriction enzymes and ligated into the prepared pET-21d vector. After the transformation of the ligation mix into DH5α' cells, colonies were screened for the positive clones. The identity and integrity of the isolated genes were confirmed by the DNA sequence analysis (ABI 377 sequencer, dye-terminators chemistry) with T7 and T7terminator primers. Recombinant plasmids were named pET-21d/**s**Mig.*Mth*II and pET-21d/Mig.*Mth*II for the truncated and full-length Mig.*Mth*II gene, respectively.

Due to the incorporation of the restriction sites, some changes in the sequences of both proteins occurred: the last amino acid "*Cys*" was removed from C-terminal part, instead, two additional amino acids – "*Leu*" and "*Glu*" - were attached, as well as six "*His*" molecules. In case of the full-length Mig.*Mth*II, the second amino acid "*Ile*" (ATT) was changed into "*Val*" (GTT). For the truncated protein, the appropriate "*Leu*" (CTT) was transformed into" *Val*" (GTT). The first codon for both proteins was set to "*Met*" (ATG). Representation of the full-length Mig.*Mth*II embedded into the pET-21d vector is in Fig 5.

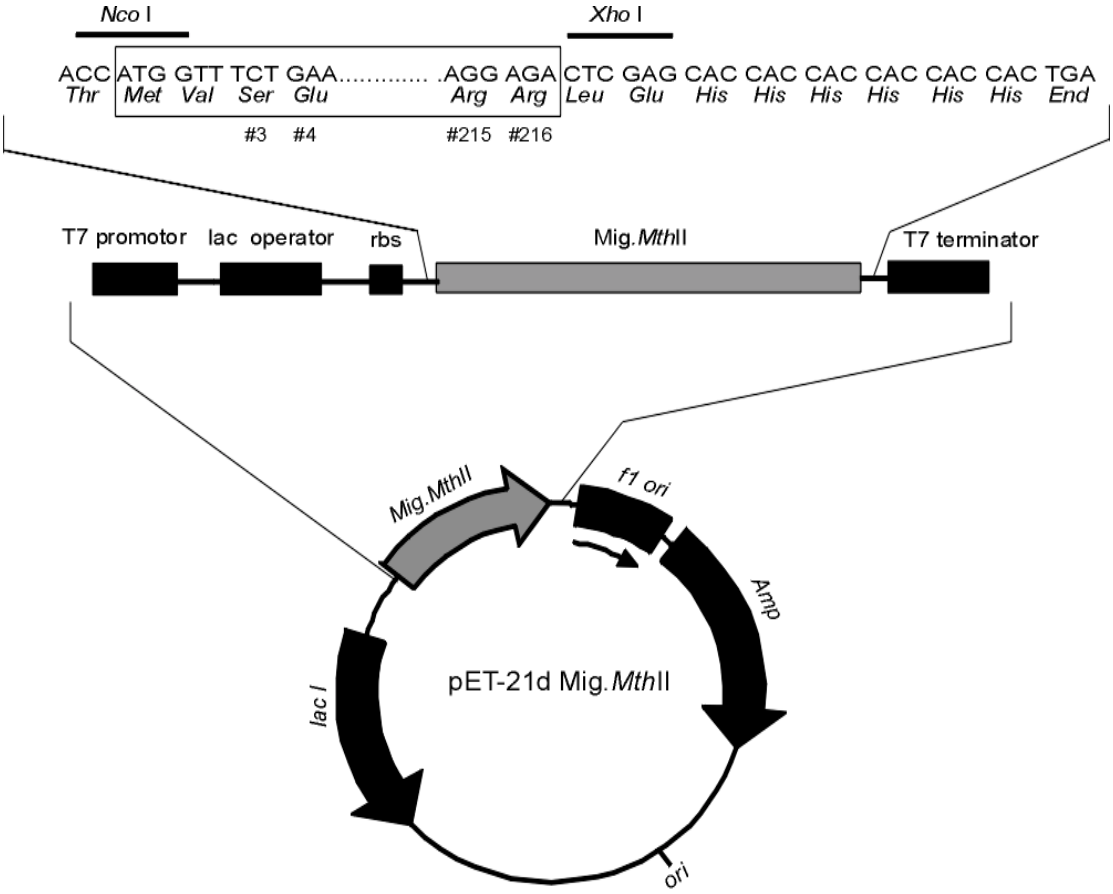


Fig 5. Embedding of the *mig2* gene in expression vector pET-21d (Studier *et.al*., 1990). Upper part: Schematic representation of Mig.*Mth*II/His-tag fusion protein. The box indicates the coding part of *mig2*. The solid line shows the position of *Nco*I and *Xho*I restriction sites.

4.1.5 Expression and purification of the full-length Mig.*Mth*II

The gene of Mig.*Mth*II was cloned under the strong T7 bacteriophage promotor, which is recognised by the T7 RNA polymerase in a highly specific manner (Tabor and Richardson, 1985). The latter enzyme is encoded by a chromosome of BL21(DE3)pLysS cells under the control of *lacUV5* promotor. To ensure that no traces of the target protein are produced in the non-induced cells, the "second control" element, the sequence of *lac* operator was placed downstream of the T7 promotor. *Lac*- repressor is encoded by the plasmid pET-21d. In the non-induced state, the *lac*-repressor binds to the *lacUV5* promotor and inhibits the synthesis of the T7 polymerase. It also binds to the *lac* sequence downstream T7 promotor and inhibits

the transcription of target gene by any of the T7 RNA polymerase made by the cell. Addition of the IPTG induces transcription of the T7 RNA polymerase, which consequently can freely access the T7 promotor (Studier *et.al*., 1990).

4.1.5.1 Expression test of the full-length Mig.*Mth*II

For the protein expression studies, the plasmid pET21/Mig.*Mth*II was transformed into BL21(DE3)pLysS. Single colony was inoculated in 50 ml 2 xYT media, containing 50 μg/ml ampicillin and 34 μg/ml chloramphenicol and incubated overnight. The cells were collected by centrifugation and resuspended in 50 ml of fresh 2 xYT media with the appropriate antibiotics. The culture was diluted 1:50 in fresh 2 xYT medium and incubated 3-4 h at 37 ^{o}C until OD_{600} reached 0.6-0.7. Synthesis of Mig.*Mth*II was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated for additional 3 h at 30 °C and 37 °C, independently.

Samples of the induced culture were collected every hour with following centrifugation, resuspention in the IMAC buffer without imidazole and sonication. Soluble and insoluble fractions were separated by centrifugation for 15 min at 15000 rpm, 4^{o}C and analysed on the 15% SDS - gel (Fig. 6).

In Fig. 6, it is clearly seen the appearance of the band between 21.5 and 31 kDa, what is in a good agreement with the calculated molecular mass of Mig.*Mth*II (26.2 kDa). Mig.*Mth*II appears already one hour after induction in the soluble fraction of the cells. Similar expression profile was obtained testing the efficiency of the protein synthesis at 30 and 37^{o}C. The amount of protein produced remained almost unchanged in the course of prolonged synthesis time (up till 6 hours). No detectable amount of the full-length Mig.*Mth*II was found in the cell pellet (data not shown).

The above described expression conditions were used for the large-scale purification of the Mig.*Mth*II.

4.1.5.2 Purification of the Mig.*Mth*II

Strain BL21(DE3)pLysS carries plasmid pLysS encoding the gene of the T7 lysozyme. The enzyme is a bifunctional protein: it is a natural inhibitor of the T7 RNA polymerase, thus helps in eliminating the background transcription level. The T7 lysozyme also acts on peptidoglycan layer of *E.coli* (Inouye *et.al*., 1973). The latter feature was used to facilitate and to complete lyses of the cells during the protein purification.

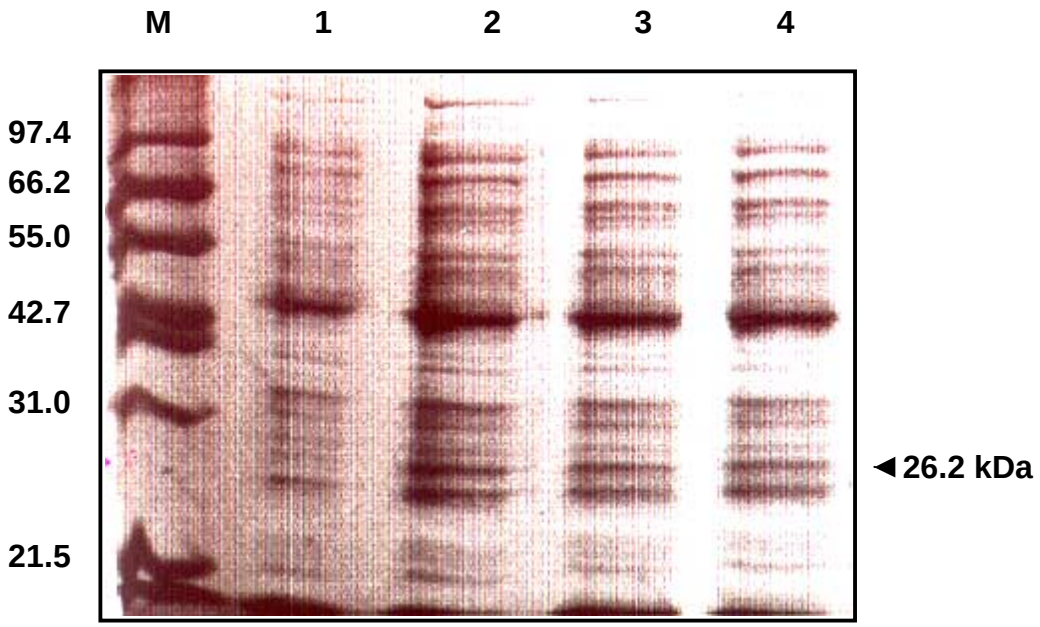


Fig 6. Test of expression of Mig.*Mth*II. **M** – Mid-range protein molecular weight marker (Promega), **1** – 10 μl of soluble fraction of crude extract before induction, **2, 3, 4** – 10 μl of soluble fraction of crude extract after 1, 2 and 3 hours after induction, respectively, 26.2 kDa – calculated molecular weight of Mig.*Mth*II.

After induction of the Mig.*Mth*II synthesis for three hours, the cells were collected by centrifugation and suspended in IMAC buffer without imidazole. Freezing of the cells at –70^{o}C and rapid thawing at 25-37^{o}C allows efficient lysis of cells by the T7 lysozyme (the culture gains cloudy and white appearance). The following sonication was used to ensure complete lyses as well as disruption of the DNA molecules.

Due to the intrinsical thermostability of Mig.*Mth*II, a heat-treatment step could be efficient way in removing most of *E.coli* proteins. Purification schemes with and without heat-treatment of the crude extract at 65 °C for 15 min were performed and compared (data not shown). Usually, after heat-treatment cleared lysate contained much less of the contaminating proteins. The slimy appearance of the culture, most probably caused by not-fully sonicated DNA, has also disappeared, thus reducing the danger of clogging the IMAC column. However, no significant difference in the final purity and yield of Mig.*Mth*II was observed, thus heat-treatment step was left optional.

After the preparative steps, cleared crude extract was loaded onto an IMAC column. The principle of the IMAC - immobilised metal affinity chromatography is based on the fact that the protein is adsorbed by coordinate bonding of the histidine residues (mainly those artificially inserted during the cloning and, in much less extend, naturally occurring ones) with nickel immobilised on the matrix. Specific elution of the protein from the column is performed by competition for nickel binding sites between the protein and imidazole (Porath *et.al.*, 1975).

Accordingly, Mig.*Mth*II was eluted from the IMAC column by increasing imidazole concentrations (Fig 7).

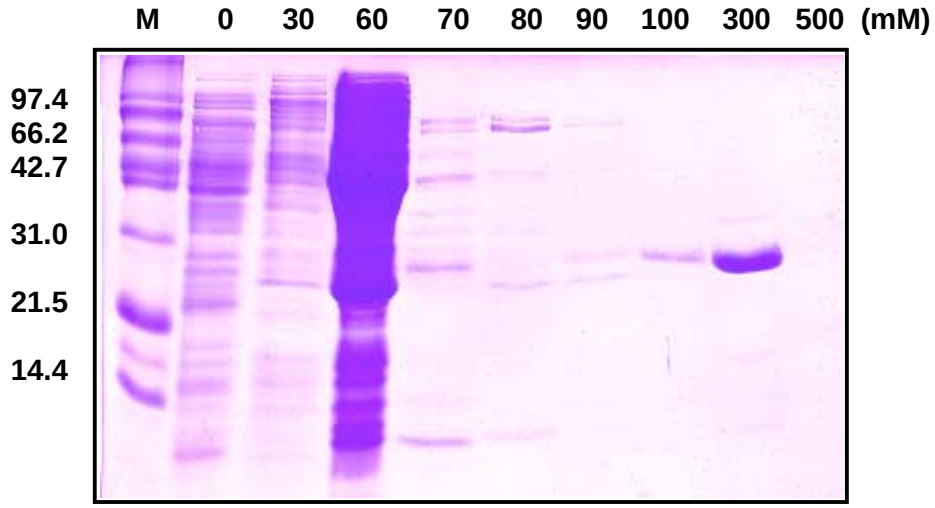


Fig 7. Purification of the Mig.*Mth*II by IMAC. **M** – Mid – range protein molecular weight marker (Promega). **mM** – 15 µl of the fractions, eluted with imidazole added at concentration indicated.

Mig.*Mth*II eluted in the range of imidazole concentration of 100 - 300 mM. The protein containing fractions were pooled and subjected to the further purification. These fractions were slowly diluted to a 1:10 ratio in a 25 mM HEPES-KOH, pH 7.6 buffer to decrease concentration of NaCl. In case of appearance of white precipitants, filtration of the protein solution through the 0.45µm filter is necessary, but this significantly reduces the yield of protein. Alternative method to remove NaCl would be dialysis against a buffer without the respective salt, but then the time for the purification procedure is significantly increased.

Further purification of the protein was done by perfusion chromatography. The method is a recent modification of adsorbtion chromatography. The medium has large "through-pores" for carry the molecules into the interior of the bead. From these pores many short "diffusive pores" branch out and, consequently provide large internal surface for the specific binding (Afeyan *et.al*., 1990). As a result, the separation of molecules can be done at high speed, resolution and big capacity. Different functional groups can be immobilised on the carrier, thus allowing the use of electrostatic, hydrophobic or affinity interactions for the protein purification. Thus, for the final purification of Mig.*Mth*II, heparin POROS column was used. Heparin is a linear and highly sulfated glycosaminoglycan able to mimic properties of DNA. Thus many DNA-binding enzymes can be separated from the pool of other proteins (Farooqui, 1980).

Elution was done by a linear 0-1.5 M NaCl gradient. The process was monitored by absorbance measurements at 260 and 280 nm. Elution profile consists of two peaks: the contamination peak appearing during the wash of the column without the salt (data not shown) and the peak at about 0.3 M NaCl concentration (Fig. 8, A). Analysis of the fractions revealed that essentially homogenous Mig.*Mth*II is present in the second peak (Fig. 8, B).

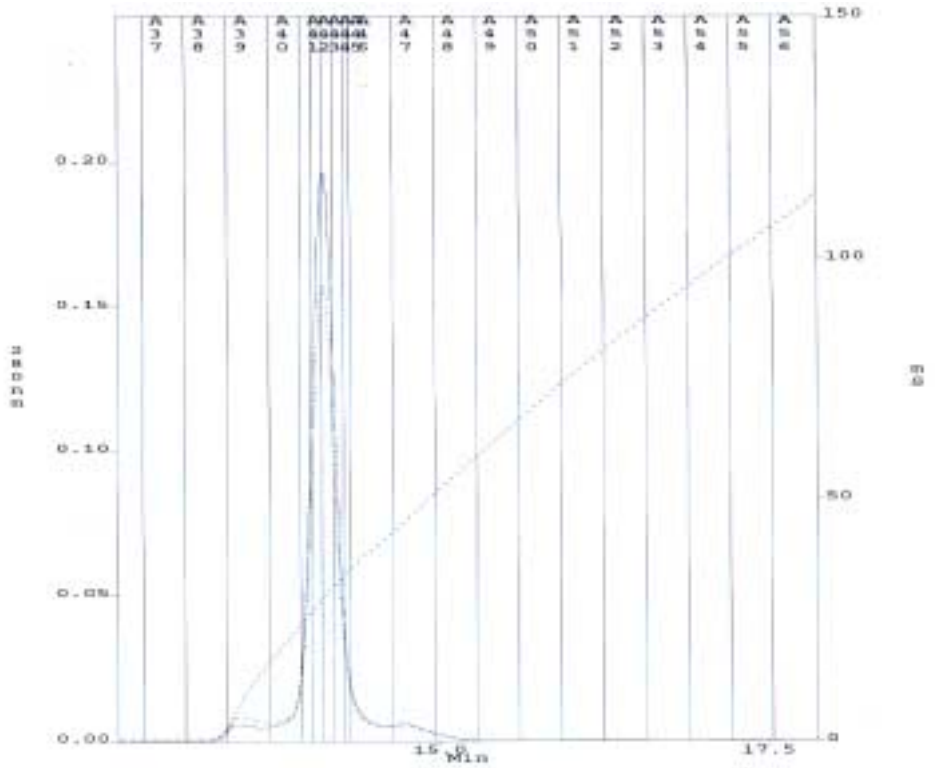
The purification procedures of Mig.*Mth*II and Mig.*Mth*I are similar, with the difference that Mig.*Mth*II elutes from the heparin column at about twice less NaCl concentration than Mig.*Mth*I.

For characterisation of Mig.*Mth*II, fractions 41-44 were combined (Fig. 8, B). Concentration of the protein and removal of NaCl was done by Centriprep filter devices with cut-off value 3000. Desalted and concentrated protein (till about 1 mg/ml) was diluted with 50% sterile glycerol and DTT added to end concentration of 1 mM. Mig.*Mth*II was stored at -20^{o}C for more than 12 months without loss of activity.

The protein expression and purification scheme described above resulted in about 2-3 mg of Mig.*Mth*II from 1l of culture.

It is known that some representatives of helix-hairpin-helix glycosylases, like endonuclease III and MutY proteins, contain [4Fe-4S] – cluster coordinated at the C-terminal part of the protein. Unique spacing of *Cys* ligands separates these glycosylases into distinct group of [4Fe-4S] – containing proteins (Golinelli *et al*., 1999) with the cluster being responsible for the enzymatic activity, proper folding and DNA binding.

A.



B.

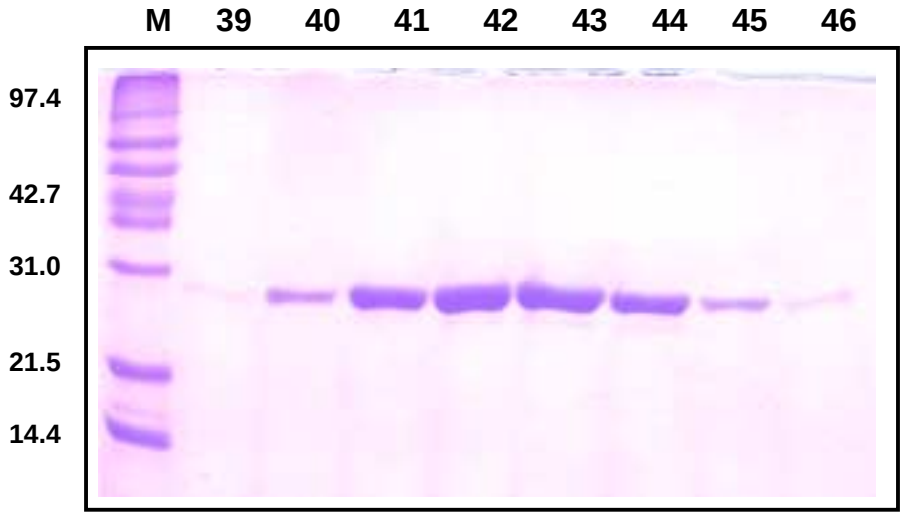


Fig. 8. Purification of the Mig.*Mth*II by heparin column (HE 20, 4.6 mm x 100 mm, 20 ml/min flow rate). **A**: Chromatograme of the run. The clear peak line represents absorption of the fractions at 280 nm, the dashed line – absorption at 260 nm. The straight dashed line is NaCl gradient, expressed in a conductivity unit milisiemens (ms). 70 ms = 1M NaCl concentration. **B**: 10 µl of peak fractions loaded, figures at the bottom line represents the numbers of the tubes, **M** - Mid – range protein molecular weight marker (Promega).

It was experimentally measured that Mig subclass glycosylases Mig.*Mth*I (Horst, 1996) and Pa-MIG (Yang *et.al*., 2000) also contain the [4Fe-4S] cluster. The latter is easily detectable by the absorption measurement at 400 nm wave-length (Sweeny, 1980) and typical yellowish – brown colour of the protein solutions.

Purified Mig.*Mth*II was also checked for the presence of [4Fe-4S] cluster by scanning the protein solution within 250 – 500 nm wavelength interval (Fig. 9).

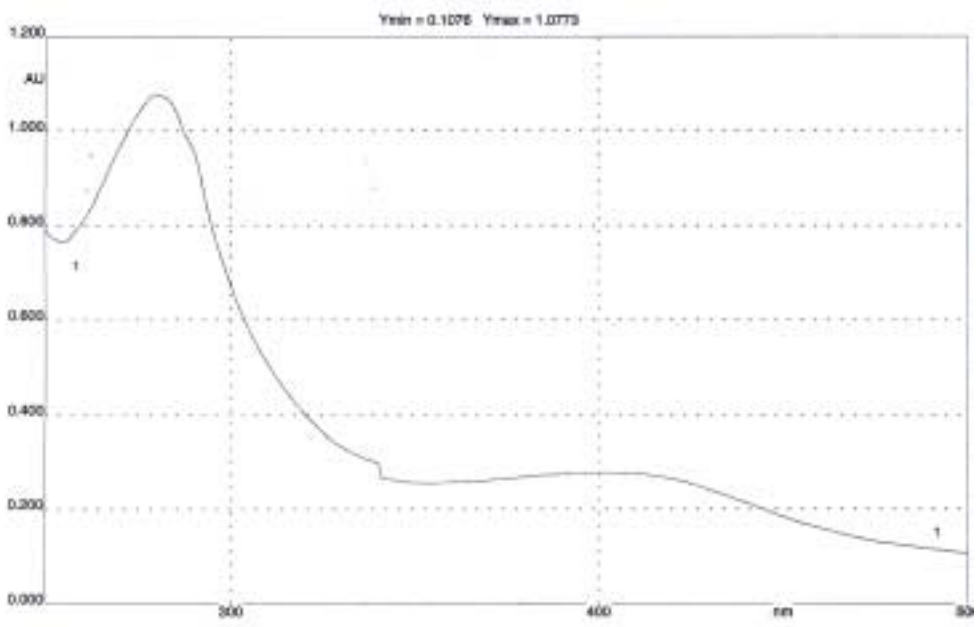


Fig. 9. Absorbtion spectrum of the Mig.*Mth*II. Ordinate represents the absorption of the protein, abscise – wavelength in nm.

Absorbtion spectrum demonstrates the presence of the additional chromophore - [4Fe-4S] group - bonded to Mig.*Mth*II, giving the protein the characteristic yellowish-brown colour.

4.1.6 Expression test of the truncated Mig.*Mth*II

The expression test similar to that described for the full-length Mig.*Mth*II was performed (Fig 10).

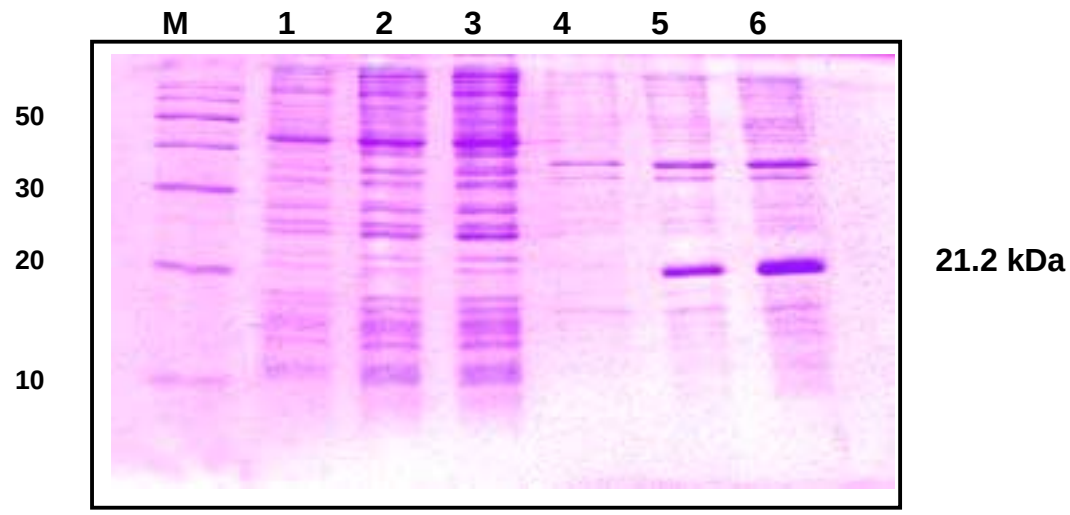


Fig. 10. Expression test of truncated version of Mig.*Mth*II. **M** – 10 kDa protein size marker (GibcoBRL). **1** – supernatant before induction, **2** – supernatant 2 h after induction, **3** – supernatant 3 h after induction, **4** – pellet before induction, **5** – pellet 2 h after induction, **6** – pellet 3 h after induction.

As it is see from Fig 10, the truncated version of Mig.*Mth*II was expressed in the clearly bigger amounts than the full-length enzyme, but most of it remained in the pellet. It is the most probable that the protein is trapped into the inclusion bodies due to the huge expression level. On the other hand, formation of the soluble protein would also be expected. To test this, the supernatant was loaded in the IMAC column. Unfortunately, no band of the respective size (21.2 kDa) was visible in the fractions after the elution with different imidazole concentrations.

It therefore seems that the first 40 amino acids are indeed needed for the proper folding and stability of the Mig.*Mth*II. Interesting to note that the cell pellet was dark brown, what indicates that not folded truncated Mig.*Mth*II is still able to coordinate [4Fe – 4S] cluster.

4.1.7 Characterisation of Mig.*Mth*II as a monofunctional U/G, T/G DNA glycosylase

Due to the high sequence similarity to other Mig subclass glycosylases (Fig. 2), Mig.*Mth*II was also expected to have the U/G, T/G cleavage activity. To test this, DNA-glycosylase assays were carried out essentially as described by Horst, 1996.

Reactions were performed with double or single stranded DNA substrates carrying fluorescein (FITC) label on the 5’ end. Concentrations of the substrates were

limited to 40-50 fmol/per single reaction due to the detection limits of the A.L.F. sequencer. Other parameters of the reaction, like concentration of the Mig.*Mth*II, incubation temperature and AP-site hydrolysis were adjusted according the protocols optimised by Fondufe, 1999.

Thus the standard reaction mixture for Mig.*Mth*II activity assay consisted of 40 fmol fluorescently labelled substrate and 40 pmol enzyme in 20 µl of 20 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$, pH 7.5 (at room temperature). The reaction was carried out at 50 °C for 60 min (then pH of reaction buffer expected to be at a value of 6.8). AP-sites were cleaved by adding 2 µl of 1 M NaOH and keeping the mixture at 95 °C for 5-10 min. For gel electrophoretic analysis, 10 µl of A.L.F. loading buffer were added and samples were loaded onto denaturing polyacrylamide gel (Pharmacia A.L.F. DNA sequencer) in 10 µl aliquots. After the separation of the fragments, the activity peaks were visualised and analysed by Fragment Manager VI.2^{TM} program (Pharmacia Biotech). For the assaying the glycosylase-associated AP lyase activity, the alkali/heat treatment was omitted. Results are summarized in the Fig. 11.

Initial activity test of the Mig.*Mth*II confirmed that the enzyme is a member of Mig glycosylase subfamily. The mismatches U/G and T/G were efficiently processed in a similar way to the other experimentally tested glycosylases: Mig.*Mth*I (Horst and Fritz, 1996) and Pa-MIG (Yang *et.al*., 2000). To distinguish the uracil cleavage activity of the Mig.*Mth*II from the possible remaining uracil-glycosylase of *E.coli*, the reactions were repeated with pre-heated enzyme (15 min, 65 ^{o}C). The results were identical to non-heated enzyme (data not shown).

The enzyme was also tested on the single stranded DNA substrate with U and T bases in the respective position. As shown in Fig. 11, no activity was detected, which is in a good agreement with published data for the other Mig enzymes.

The experiment also showed that Mig.*Mth*II glycosylase like Mig.*Mth*I is a monofunctional enzyme, since no product formation was detected after omission of the alkali-heat treatment step. The finding is consistent with the predicted mechanism of AP-site cleavage by glycosylase-associated lyase activity (Dodson *et.al*., 1994). It is believed, that the enzyme would possess bifunctionality, if an attack on glycosidic bond is initiated by nucleophilic *Lys* residue, like K120 in Endo III (Thayer *et.al*., 1995). Consequently, the respective position among monofunctional enzymes would be expected to contain different residues. Frequent substituents are serine or tyrosine with the latter amino acid present in the sequences of both monofunctional Mig.*Mth* gylcosylases (Y126 for Mig.*Mth*I and Y124 for Mig.*Mth*II).

Closely related Pa-MIG glycosylase, though containing *Tyr* in the respective position, is able to process AP-site with significant activity (Yang et.al., 2000). Similar can be said about MutY of *E. coli*, which exposes little, but detectable lyase

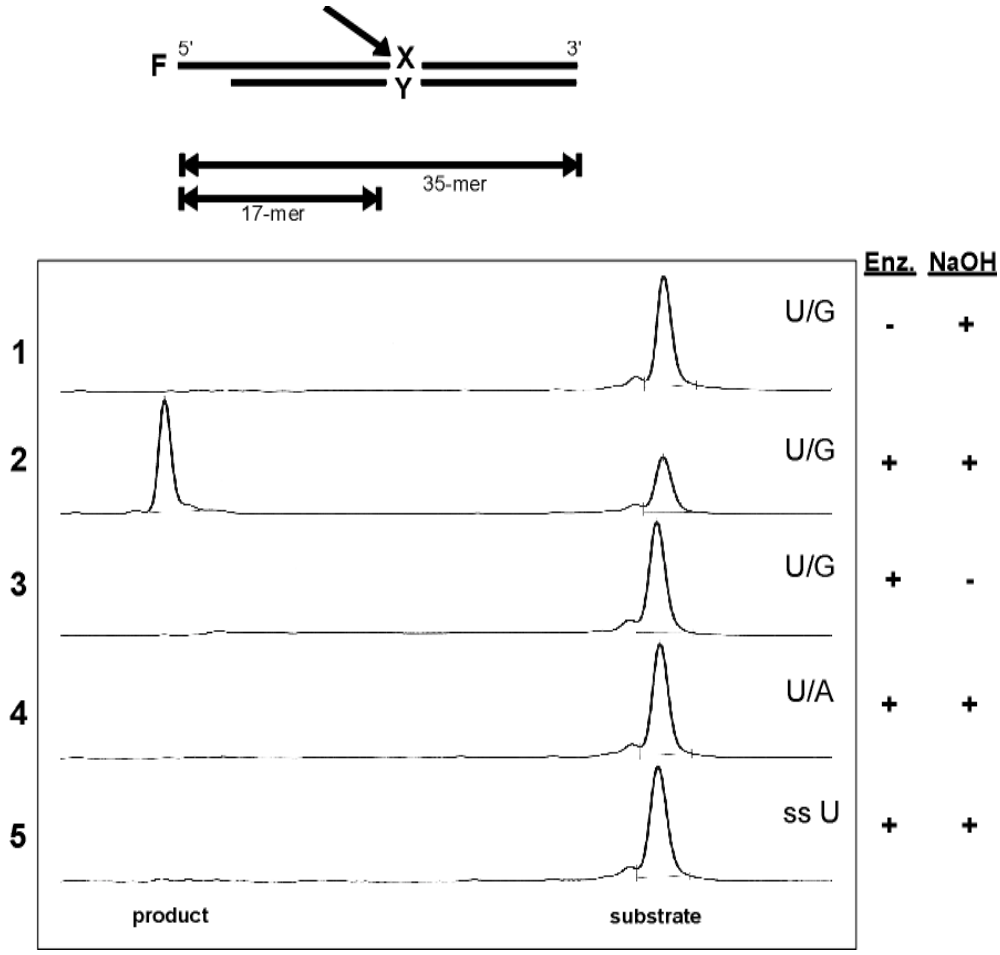


Fig. 11. Reactions catalysed by Mig.*Mth*II. Synthetic DNA duplexes used in enzyme assays are illustrated at the top. F = fluorescein moiety, X = U or T, Y = G or A; the arrow points to the site of cleavage upon glycosylytic removal of base X, followed by the alkali/heat-induced strand scission (the length of fluorescently labelled strand of substrate and product indicated). ssDNA: as above, but the upper strand only. The lower part documents the gel electrophoretic analysis of DNA fragments. Assay conditions are summarised on the right margin: “plus” and “minus” signs for addition/no addition of the enzyme and, respectively for treatment with the alkali/heat. Analogous test with T instead of U as a base X yields a qualitatively identical picture (data not shown).

activity, but lacks the lysine residue in the expected position (Scott and Sheila, 1998). However, both observations could possibly arise from experimental artefacts due to

spontaneous hydrolysis of AP site, which is particularly increased at high temperatures.

Also, similar to other Mig glycosylases, no activity on U/A mismatch was detected.

4.1.8 Qualitative comparison of Mig.*Mth*II and Mig.*Mth*I

The initial test of enzymatic activity of the Mig.*Mth*II described the enzyme as a double stranded specific, U/G, T/G mismatches processing monofunctional glycosylase. The same characteristics are known for the Mig.*Mth*I, isolated from the plasmid of *M. thermoautotrophicum* THF (Horst and Fritz, 1996).

4.1.8.1 Removal of U/G and T/G mismatches from different sequence contexts by Mig.*Mth*II and Mig.*Mth*I

From the work of Fondufe, 1999, it is known that Mig.*Mth*I has a modest rate of sensitivity to the flanking sequence around the mismatch. Also, no discrimination between U/G and T/G base/base oppositions was detected.

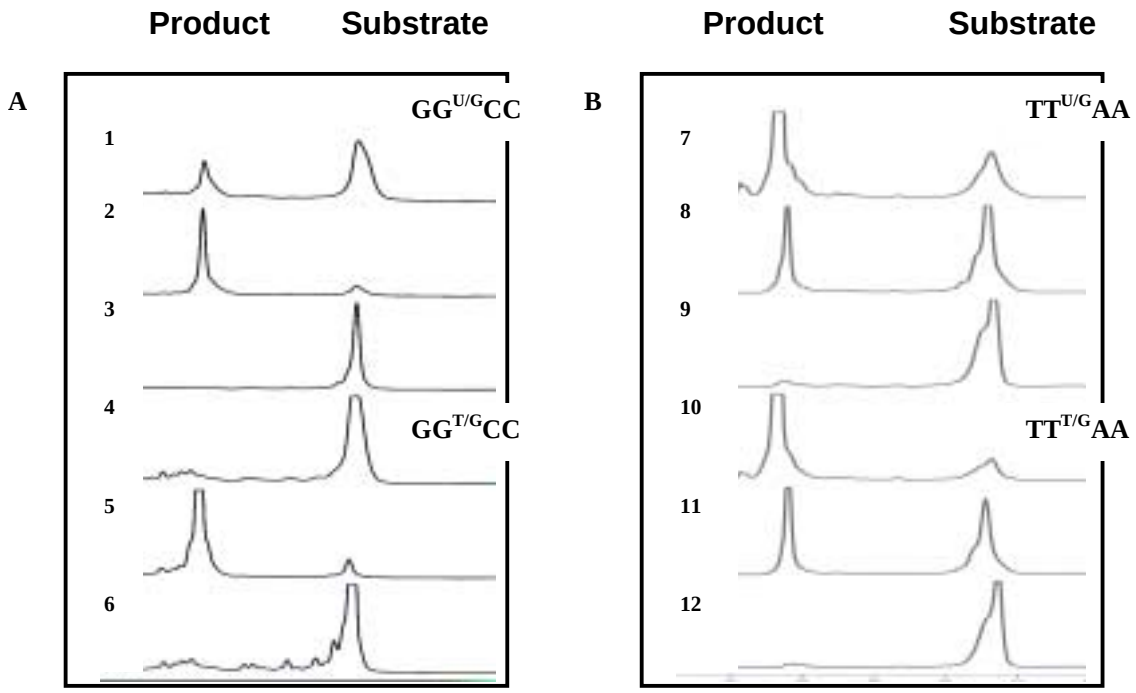


Fig. 12.. Cleavage efficiency of the U/G and T/G mismatches by the Mig.*Mth*II in the different sequence contexts. **1, 4, 7**, and **10** – Mig.*Mth*II, **2, 5, 8**, and **11** – Mig,*Mth*I, **3, 6, 9** and **12** – no enzyme included. All the reactions were carried out at 55 ^{o}C, 1h.

The best and the worst substrates of the Mig.*Mth*I were taken for the initial test of the Mig.*Mth*II. Reactions with both enzymes were performed in parallel under the identical conditions. Results of A.L.F. run are shown in the Fig. 12. This first experiment revealed striking differences. One of the substrates, $TT^{X}AA$, was processed by both enzymes with the comparable efficiency. The another one, $GG^{X}CC$, was fully cleaved by the Mig.*Mth*I, but only marginally by the Mig.*Mth*II. Even more, only U/G mismatch was processed, while the T/G remained uncleaved. This is different from Mig.*Mth*I, which cleaves both base/base oppositions, irrespective of the sequence context.

Thus, otherwise similar enzymes Mig.*Mth* glycosylases purified from different strains of *M. thermoautotrophicum*, bear significant differences towards the type of the mismatch as well as the sequence context surrounding this mismatch.

4.1.8.2 Test of substrate spectrum of the Mig.*Mth*II

Further comparison of the properties of Mig.*Mth* enzymes was done by qualitative test on the T/G mismatch embedded into 20 various sequence contexts. Some of the duplexes had been used for the earlier characterization of the Mig.*Mth*I (Fondufe, 1999). Their variable sequences were derived from the TIM methyltransferase recognition site (Nölling and de Vos, 1992). Use of these substrates served for the direct estimation of the differences between the Mig.*Mth*II and Mig.*Mth*I. Other substrates were provided by Dr. M. Laging (Laging, 2000) and contained the patterns different from recognition sequence of TIM methyltransferase. The reactions were performed for 30 min at 50^{o}C. The cleavage efficiency was expressed as a part of reduction in fluorescence of substrate, when the product and substrate fluorescence values taken together as 100%. According to the percent of substrate consumption, four groups of the sequences were separated: well cleaved, intermediate, slowly and non-cleaved (Table 9).

The experiment revealed several important substrate recognition features of Mig.*Mth*II. First, numerous non-cleaved sequence patterns were found. This observation was never described for any of DNA repair enzymes. Second, it became obvious that the Mig enzymes possess reversed sensitivity to the nature of flanking bases

Table 9 shows that Mig.*Mth*II does not cleave substrates containing GG or GA (stretches of purines) on the 5' side of the mismatch and CC or C coupled with any other base, but not G, on the opposite side. On contrary, the substrates cleaved with the highest efficiency require CC in front and GG behind the mismatch. Clearly, the

Efficiency of cleavage	Sequence context
Well – cleaved substrate (80% - 100%)	ccg**C^{T}CGGG**ttaa ccg**CC^{T}GGG**ttaa
Intermediate substrate (30% - 70%)	ccga**TT^{T}AA**ttaa ccg**TA^{T}TAG**ttaa ccg**A^{T}CGGT**ttaa ccg**AC^{T}GGT**ttaa ccg**TC^{T}CTA**ttaa
Slowly – cleaved substrate (10% - 20%)	ccga**GA^{T}GC**ttaa ccga**GA^{T}AC**ttaa ccga**GC^{T}GC**ttaa ccga**GT^{T}GC**ttaa ccga**TG^{T}CT**ttaa ccg**G^{T}GCGC**ttaa ccg**GCG^{T}GC**ttaa
Not – cleaved substrate*	ccga**GG^{T}CC**ttaa ccga**GG^{T}CA**ttaa ccga**GA^{T}CC**ttaa ccga**GA^{T}TC**ttaa ccg**A^{T}CCCT**ttaa ccg**A^{T}CAAT**ttaa

Table 9. Initial estimation of sequence context influence upon T/G mismatches cleavage by Mig.*Mth*II. * - substrates which were not processed also in the prolonged time of the incubation. All substrates were cleaved by the Mig.M*th*I (data not shown).

enzyme is sensitive to the polarity of G and C bases around the mismatch. Mig.*Mth*I, on the other hand, prefers an inverted situation: the best-cleaved substrate has GG on the 5' side of the mismatch and CC on the 3' side ((substrate GG^{X}CC) Fondufe, 1999), while CC^{T}GG is cleaved with significant delay (quantitative data will be shown later).

The group of slowly cleaved substrates by Mig.*Mth*II, thus, contains purines on the 5' side of the mismatch. However, the appearance of pyrimidines on the 5' and purines on the 3' side of the mismatch facilitate cleavage (e.g., GA^{T}TC is not cleaved, GA^{T}GC is readily processed). Again, the reverse picture was obtained for the Mig.*Mth*I gylcosylase: exchange of purines into pyrimidines on the 5' side and *vice versa* on the 3' side, retarded the activity of the enzyme (e.g., GA^{T}TC is an intermediate substrate, while GA^{T}GC is one of the worst ones).

These observations correlate well with the detailed substrate spectrum analysis performed for Mig.*Mth*I (Fondufe, 1999). The analysis was based on the assumption that the enzyme should work best in the sequence context derived from the

recognition site of the respective methyltransferase (Horst, 1996). Indeed, this turned to be true for the Mig.*Mth*I. The sequences most resembling the GGXCC patterns were cleaved more efficiently (Table 13, Fondufe, 1999). Consequently, the distribution of purines and pyrimidines around the mismatch is an important factor in the substrate recognition process by the Mig.*Mth*I.

Similarly, the optimal substrate for Mig.*Mth*II should also be derived from methylation pattern of respective methyltransferase, since it is also found close to the gene of the methyltransferase (*orf* 495). Unfortunately, nothing is known about the DNA methylation activities in *M. thermoautotrophicum* ΔH. Therefore, the substrate, which was experimentally tested to be the most efficiently cleaved, was referred to be as the "optimal" substrate for Mig.*Mth*II.

4.1.9 Optimisation of A.L.F run conditions

The experiments demonstrated above showed unusual selectivity of the Mig.*Mth*II enzyme to various DNA substrates. To evaluate this selectivity, quantitative studies were performed.

However the initial runs on A.L.F. showed an unusual performance (Fig. 13, C). When two or more substrates were incubated with the enzyme and applied to gel, the substrate peaks were clearly seen, but only one peak in the position of product migration. This could be due to several reasons: a) one of the substrates is not cleaved; b) product peaks run at the same time implying the same length. To solve this, substrates were tested by separate cleavage, which showed normal performance. On the closer look, the mistake was found in the design of oligos. Usually, substrates were designed so that difference among several full – length substrates and, consequently, their products was kept constant – 4 nucleotides (Fondufe, 1999). This was achieved by maintaining the position of mismatched nucleotide constant and extending the sequence into 5' end by addition of the bases. However, in some of the substrates the position of mismatch nucleotide was changed (Laging, 2000). Thus, the difference between the product peaks appeared to be less than 4 bases and could not be resolved under the conditions used before (Horst, 1996). This explained, why several products migrated as the peak (Fig. 13, C). The problem was solved by increasing the resolution power of polyacrylamide matrix. For this, the gel density was increased from 6% to 10%. Simultaneous decrease of the migration speed of DNA through the gel from 34W to 25W, resulted in a clear separation of the fragments, those were different only in two bases (Fig. 13, A).

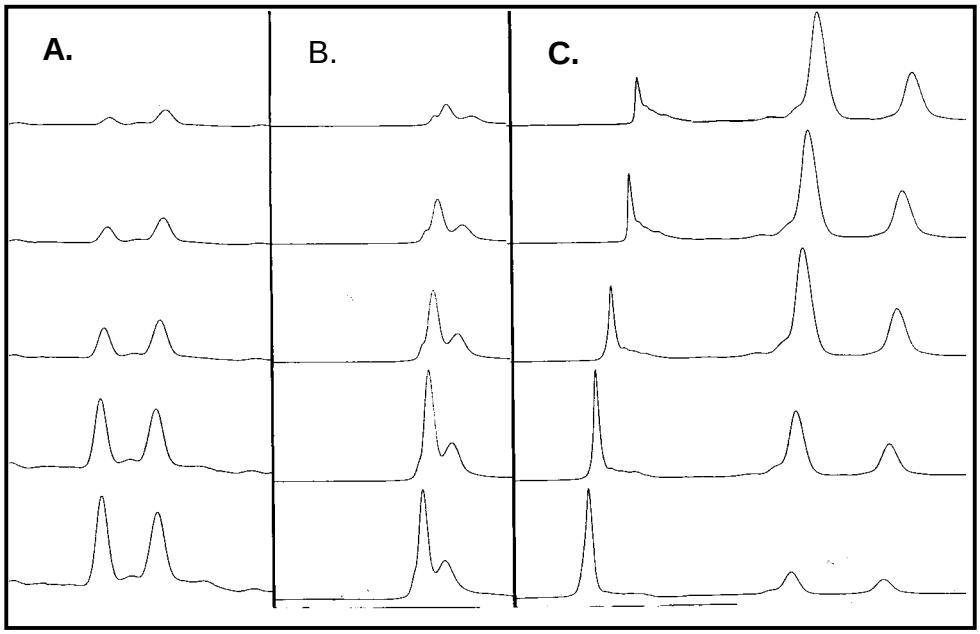


Fig. 13. Optimization of the A.L.F. run conditions. **A**: 10 % acrylamide:bisacrylamide, 50 ^{o}C, 25W, 60 mA, 1500 V, **B**: 8,5% LongRanger, 50 ^{o}C, 25W, 60 mA, 1500 V, **C**: 6% acrylamide:bisacrylamide, 37 ^{o}C, 34W, 38 mA, 12400 V.

Besides this, two different polymer matrixes were tested. One of them, LongRanger^{TM} (FMC BioProducts), is known to build very stable polymer, which is not degraded even after long runs. However, with this, only partial separation of DNA fragments was achieved (Fig. 13, B). For this work, rigid matrix formed by 10% acrylamide was the most suitable (Fig. 13, A). Additionally, appearance of a more symmetrical peak was obtained by raising the running temperature of the gel from 37 ^{o}C to 50 ^{o}C.

These changes significantly extended the time needed for the enzymatic activity analysis, but enabled separation of DNA fragments, that differ by only two bases. This opened the possibility of using more substrates for a single reaction and performing the multiple substrate tests in a more efficient way. Clear and symmetrical shape of the peak made integration of the data easier. Finally, “dirt peak” which results from the impurities in the gel and/or reaction was well resolved in time and did not interfere with the activity peaks.

4.1.10 Quantitative evaluation of the selectivity of Mig.*Mth*II

Previous qualitative experiments unambiguously revealed that Mig.*Mth*II is more selective enzyme than Mig.*Mth*I. Knowledge of the discrimination rates of cleavage of Mig.*Mth*II is important to characterise the enzyme in details and this will lead to a full understanding of its action *in vivo*. Quantitative data would also allow the direct comparison to Mig.*Mth*I glycosylase, discrimination spectrum of which is known from the work of Fondufe, 1999.

4.1.10.1 Application of multiple substrate kinetics to Mig.*Mth*II

Reaction conditions for the multiple substrate kinetics were essentially as described in **4.1.7** with the difference that several fluorescently labelled substrates were cleaved simultaneously. The assay was done in 12-fold larger volume with gross amounts of assay components scaled up accordingly. The reactions were performed at 50 ^{o}C temperature to slow down the reaction rate and to make measurements more accurate as well as to avoid the denaturing of DNA duplexes. Aliquots of 20 µl were removed within 1h at fixed time points and AP-sites cleaved by alkali/heat-treatment. DNA fragments were separated under the conditions, which are described in **4.1.9**. Areas of the peaks (Gläsner *et.al*., 1992) were integrated with the help of FragmentManager V1.2^{TM} software (Pharmacia). Remaining fluorescence for each substrate was expressed as a percentage of the overall substrate and product fluorescence and compared with that obtained for the reference substrate invariably included in all measurements. Resulting value was expressed as a relative rate constant. All necessary mathematical transformations were performed with the SigmaPlot^{TM} program. Assay for the every test substrate was repeated at least three times and averaged.

4.1.10.2 Selectivity of the Mig.*Mth*II

Specificity measurements of Mig.*Mth*II would address two questions: first, the rates of discrimination towards various sequence contexts and secondly, the difference in the cleavage of U/G and T/G mismatches within the same sequence context. For this, 4 different substrates with varying cleavage efficiencies were used (Table 9). According to Schellenberger *et.al*., 1993, the measurements of mixes of substrates with large differences of their rate constants led to significant errors in calculations. To avoid this, the reference sequence (GA^{T}CGGT) was chosen from the group of the

substrates with the intermediate reactivity (Table 9) and was invariable included in all reactions.

The analogous experiment was done with the Mig.*Mth*I under identical reaction conditions and protein concentrations. Results are shown in the Fig. 14. A summary of the relative rate constants is listed in the Table 10.

The experiment enabled the quantitative evaluation of the high selectivity of the Mig.*Mth*II. The estimated rate of discrimination for the repair of T/G mismatch between worst and best-cleaved substrates is more than 450. Moreover, the actual rate is indefinite, because in $GG^{T}CC$ sequence context T/G mismatch is not cleaved (lowest right panel in Fig. 14). The selectivity for the sequence context in case of U/G oppositions seems to be less pronounced (discrimination factor of about 80 within the range of substrates measured).

Invariably, Mig.*Mth*II prefers U/G to T/G mismatches with the discrimination factors varying from 2 – 3 for the most efficiently processed substrate, to more than 12 for the worst substrate. In some of the sequence contexts ($GG^{X}CC$, for example) the enzyme works only on uracil and does not cleave substrate with thymine within the same sequence context. Thus, the rate of discrimination between U/G and T/G mismatches is also indefinite in the “bad” substrate.

From work of Fondufe, 1999, it is known that Mig.*Mth*I is a moderately discriminative enzyme within the set of substrates tested. Difference in the cleavage of T/G between the best and worst substrates is only about 10 times. My measurements resulted in the rate of 5 to 6 for T/G removal and about 10 times for U/G removal, which is in a good agreement with the earlier data. Important to note, the enzyme was able to cleave all the substrates, even then the surrounding sequence did not resemble GGXCC pattern.

Similar to Mig.*Mth*II, the enzyme prefers U/G to T/G mismatches with discrimination factor 2 for the best cleaved substrate and no difference for the slow substrate (left side of the Fig. 14). However, when compared to Mig.*Mth*II, the order of this preference is different: Mig.*Mth*II discriminates more efficiently in the unfavourable environment ($TG^{X}CT$), the Mig.*Mth*I – within the best-cleaved substrates ($CC^{X}CC$) (Table 10).

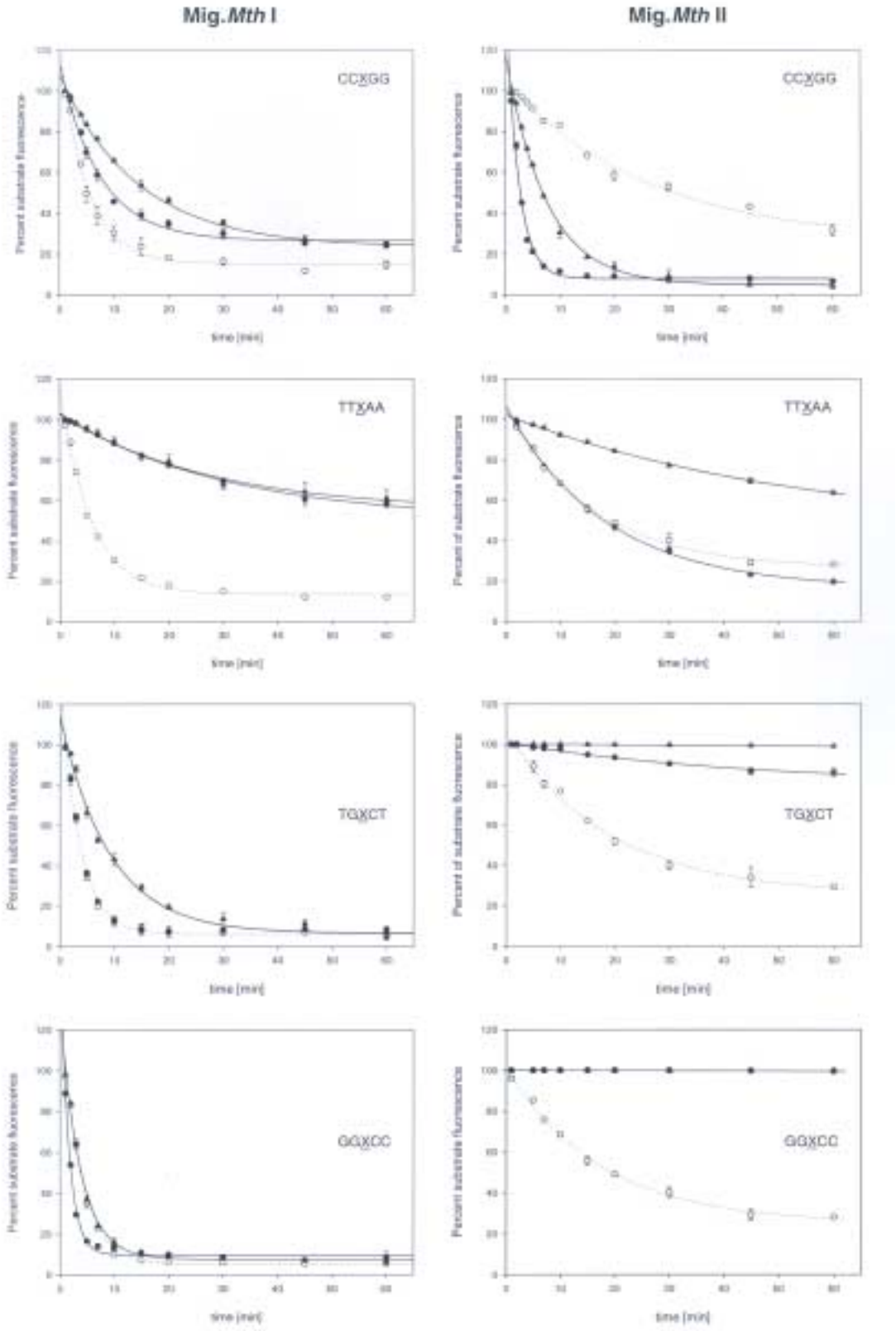


Fig. 14. Representation of multiple substrate kinetics of Mig.*Mth*II and Mig.*Mth*I. Efficiency of substrate cleavage decreases from top to bottom for each enzyme. Underlined X symbolizes U or T opposite G. Measurements with T/G mismatches, embedded in different sequence contexts are showed as black triangles, with U/G - as black circles. The white circles indicate reference substrate "GA$^{T/G}$CGGT" (for details see Materials and Methods).

	Mig.*Mth*II			Mig.*Mth*I		
Substrate	**Rank of substrate**	**Cleavage efficiency**	**U/G : T/G ratio**	**Rank of substrate**	**Cleavage efficiency**	**U/G : T/G ratio**
CC**U**GG CC**T**GG	1	9.94+/-0.15 4.65+/-0.38	2.1 : 1	3	0.65+/-0.03 0.56+/-0.06	1.2 : 1
TT**U**AA TT**T**AA	2	1.17+/-0.04 0.31+/-0.03	3.8 : 1	4	0.21+/-0.03 0.20+/-0.03	1.0 : 1
TG**U**CT TG**T**CT	3	0.12+/-0.01 0.01+/-0.01	12.0 : 1	2	0.95+/-0.02 0.49+/-0.06	1.9 : 1
GG**U**CC GG**T**CC	4	marginal cleavage no cleavage	–	1	1.83+/-0.51 0.91+/-0.04	2.0 : 1

Table 10. Comparison of the substrate specificity between Mig.*Mth*II and Mig.*Mth*I. The relative rate constants are derived from the experiment showed in **Fig. 14.**

4.1.11 Biological role of the Mig.*Mth*II glycosylase

The experiment described above revealed several important features of Mig glycosylases from *M.thermoautotrophicum* as well as their possible biological role.

To date several types of T/G, U/G mismatches specific enzymes, like Vsr.*Eco* (Hennecke *et.al*., 1991), Mig.*Mth*I (Horst and Fritz, 1996) and TDG (Neddermann *et.al*., 1996) were described. All proteins process both U/G and T/G mismatches, but the latter is considered to be a primary substrate. Consequently, repair of damages produced by DNA methylation was attributed as the main function of these enzymes. Several observations speak in favor to latter suggestion. First, Vsr endonuclease and Mig glycosylase are located in close physical proximity to C5 - cytosine specific methyltransferase (Dar and Bhagwat, 1993, Nölling *et.al*., 1992. respectively). Despite that this is not the case for TDG glycosylase, the latter is co-expressed with DNA methyltrasferase (Niederreither *et.al*., 1998). Second, the highest activity of mismatch repair for all enzymes was observed within the sequence context derived from a methylation pattern of the respective methyltransferase (Vsr.*Eco* - Gläsner

et.al., 1995, TDG – Ullah *et.al*., 1996, Mig.*Mth*I - Fondufe, 1999). In addition, T/G preference over U/G was observed for Vsr endonuclease (Laging, 2000). However, glycosylases remove uracil more readily. But this was attributed to the specific features of the active pocket of the enzymes, and repair of uracil is considered as a backup function.

Due to the close proximity to the *orf* 495, coding for the methyltransferase gene (Fig. 4), Mig.*Mth*II was also expected to possess the ability to discriminate among various DNA sequences. The results revealed surprisingly high discrimination rate (more than 450 times), that is significantly more than what has been reported to date: Vsr.*Eco* - 20 times (Gläsner *et.al*., 1995), Vsr.*Bst* - 5 - 10 times (Laging, 2000), Mig.*Mth*I - 10 times(Fondufe, 1999), human TDG - 12 times (Ullah *et.al*., 1996). This level of the discrimination is only approximate, because no “optimal” substrate for the Mig.*Mth*II is known. Thus, there possibly exist sequences, which can be processed even faster, with even higher rates.

Remarkably, some sequences (Table 10) are not processed at all. Such a high stringency to the particular sequences is described for the first time among proteins involved into DNA repair. This high stringency of Mig.*Mth*II, indicates that the enzyme possibly follows the specificity of the respective R/M – system. Consequently, the primary role of Mig.*Mth*II would be in the avoidance of genetic damages resulting from the hydrolytic deamination of DNA 5-meC residues. This enzyme could also be acting as a backup activity removing mutagenic uracil residues.

Mig.*Mth*I, on contrary, is able to repair T/G and U/G mismatches in all sequence contexts with the moderate degree of discrimination. The sequences with no or only negligible similarity to the “optimal” pattern are also cleaved efficiently. This is rather surprising, due to the close proximity of the methyltransferase gene on the plasmid. Little discrimination rates between U/G and T/G mismatches, in addition, led to the working hypothesis, that Mig.*Mth*I glycosylase could serve or even be a main player in a more general pathway of DNA deamination. The suggestion is especially genuine in the light of the fact that no other U/G and T/G cleavage enzymes have been found in the genome of *M. thermoautotrophicum*.

The experimental data revealed significant differences between the enzymatic properties of the Mig glycosylases purified from different strains of *M.thermoautotrophicum*. Thus, despite the high similarity on the level of the primary composition and unifying specificity towards U/G and T/G mismatches, it is very likely that Mig.*Mth*II and Mig.*Mth*I could participate in the distinct DNA repair pathways *in vivo*.

The qualitative measurements already showed that Mig glycosylases possess reversed specificity to the sequence context, namely in the polar distribution of purine/pyrimidine nucleotides. The quantitative results confirmed this observation (Table 10). Even more, the best substrate of Mig.*Mth*I is the worst for Mig.*Mth*II and *vice versa*. Other substrates are arranged accordingly. The latter observation is highly significant biologically. The presence of several Mig-type glycosylases with the directly inverted specificity within the same organism, could ensure sufficient protection system of the genome against the damages of hydrolytic deamination, and, consequently, solve the puzzle of the absence of other U/G removal activities.

The differences between Mig.*Mth*II and Mig*Mth*I glycosylases are not only limited by the above features. The observed reversibility of the substrate spectrum cannot be applied to the all substrates. The exception is the AT-rich substrates (Table 10). Earlier it was suggested that AT-rich region produces local melting around the mismatch and this decreases the activity of the double stand-specific enzymes (Jones *et.al*., 1987). This is very much true for the Mig.*Mth*I (Fondufe, 1999 and my measurements). However, in case of Mig.*Mth*II, AT-rich substrates are among the best substrates tested. Knowing that the enzyme is also a double-strand specific glycosylase and all the reactions were performed under identical conditions for both Migs, the influence of the local DNA melting seems to be unlikely. Or at least, constrains on the active centre of Mig.*Mth*II due to the selective interaction with the respective DNA bases overcomes the effect of the local DNA melting.

4.1.12 Possible determinants of the Mig.*Mth*II selectivity

Thus, the kinetics data of the Mig.*Mth*II unambiguously defined the biological role of the enzyme in the cell. However, the detailed explanation of the observed effects on the molecular basis is a difficult task.

It is known that the majority of uracil-specific glycosylases possesses the effective mechanism of discrimination between uracil and thymine. In most cases the thymine is excluded by the sterical hindering of the C5 methyl group (Savva *et.al*., 1995). Due to such constrains, the enzyme is not able to accommodate the thymine in the active pocket (Mol *et.al*., 1995). This seems to apply to other T/G, U/G specific enzymes, though a less strict mechanism (MUG, for example), which still allows the accommodation and removal of the thymine residue with the measurable rate (Schärer and Jiricny, 2001).

The prevalence of the uracil against thymine for the Mig class enzymes has also been documented (Yang *et.al*., 2000; this work). Most likely, the explanation also implies the sterical hindering of the methyl group of the C5 position of the thymine. Unfortunately, the three-dimensional structure of the Mig glycosylases is not available, but the approximation of the residues involved can be done due to the high similarity to the other helix-hairpin-helix class enzymes with the crystal structures available (Thayer *et.al*., 1995, Guan *et.al*., 1998).

Thus, as an interfering group the hydrophobic *Ala* 50 residue located within α2-α3 helixes was chosen (Fondufe-Mittendorf, personal communication). Exchange of this residue into the *Val*, decreased the cleavage rate of the thymine about two-fold by the Mig.*Mth*I. The natural occurrence of *Val* residue in the respective position in the sequence of PA.Mig serves as an additional proof. This correlated well with the observed prevalence towards the uracil residues of this enzyme (Yang *et.al*., 2000). Yet, the Mig.*Mth*II might use a different mechanism. The enzyme is naturally more discriminative towards thymine base (Table 10), despite that, the respective position is occupied by the *Ala* 49 residue (Fig. 2). Alternatively, the modelled active pocket of the enzyme is more geometrically constrained and even the small side chain of the alanine is sufficient to hinder the methyl group of the mismatched thymine.

Important to note, that the rate of discrimination between the uracil and thymine by Mig.*Mth*II increases in the less favourable sequence. Possibly, the binding to the “bad” substrate reduces the inner space of the active pocket so that, the thymine is fully excluded. Confirmation to this suggestion comes from the gradual decrease in the thymine removal in correlation with the decrease in the efficiency of the substrate cleavage (Table 10). This was also found for the human TDG enzyme (Ullah *et.al*., 1996). The observation was explained by the cooperative interactions not only with the mismatched base, but also with the neighbouring bases on the 5’ side. The latter bonding obviously are crucial for the correct orientation and scaffold of the active site of the TDG glycosylase (Waters and Swann, 2000).

Unfortunately, the accurate mode of the interactions between the mismatched bases and their close surrounding are not known for the Mig-type glycosylases. Thus, the deduction of the reasons for the substrate spectrum and selectivity of Mig glycosylases is currently problematic. Nevertheless, some of the answer can be derived from the data of the multiple sequence alignment (Fig. 2). Undoubtfully, the different specificity and selectivity of the Mig glycosylases is coded in the primary composition of the enzymes. The most reasonable expectation is that the region strongly divergent among the Mig glycosylases could serve as a respective sensor. Indeed, the string of about 6-7 amino acids in the approximate position of 160-170

a.a. were found (Fig. 2). According to the crystal structure of the MutY (Guan *et.al*., 1998), the respective region lies in the position of the α8-α9 helixes, which comprises the [4Fe-4S] domain. Despite that the region is not in the direct interaction with the predicted position of the active pocket, but could scan for the particular base pair edges, thereby facilitating and/or precluding the flipping of the mismatched base.

Remarkably, the conserved motif "DSSRRS" within this region is present only in the sequences clustering around Mig.*Mth*II (Fig. 2). These glycosylases are not only highly conserved, but also their closest genomic loci are well preserved. One of the features of their surrounding is the presence of the gene of the putative methyltransferases (Fig. 24). From the protein sequences of these methyltransferases, the similar methylation pattern is expected and, consequently, the similar sequence specificity of the glycosylases (see **4.4.2**). Thus, the presence of the unifying motif "DSSRRS" could serve as a reasonable determinant of the shared specificity. The appropriate sequence in the Mig.*Mth*I ("NLNYNHK") is completely divergent, which is in a good agreement with the profound differences in the kinetics data between the enzymes (Table 10). The Mig's from the extrathermophilic *A. pernyx* and *P. aerophylum* share the motif "KKRPHTD" in the relevant position. Thus, the similar specificity is expected. In confirmation to that, the sequence data of their neighbouring methyltrasferases indicates the presence of a similar methylation pattern (**4.4.2**).

Surprisingly, the possible variable motif is very short. Other DNA sequence specific proteins contain significantly longer patterns. For example, variable region of the DNA methyltransferases is about 150 a.a. long and the minimal size of the target recognition domain TDR is approximately 40 a.a. (Cheng and Blumenthal, 1999). The restriction endonucleases are non-conserved through the whole length of the protein (Lauster *et.al*., 1989) and even can not be found by the similarity search. However, the detailed mechanism of the interaction between the Mig glycosylases and DNA substrate molecules is not known. Likely, the residues, involved in the specific interaction with DNA, are spread rather concentrated in the particular site. Therefore, more detailed analysis is needed, particularly in the light of the fact that the sequence specificity of helix-hairpin-helix subfamily has never been addressed.

4.1.13 Direct comparison of the U/G and T/G cleavage by the Mig.*Mth*II and Mig.*Mth*I

The difference in the efficiency of the cleavage of the T/G and U/G mismatches is quite small (Table 10)**,** particular by the Mig.*Mth*I enzyme. To ensure the accuracy of the kinetics assays and to avoid any unexpected shift in the results due to the inclusion of the reference sequence **(4.1.10.1)**, the additional independent experiment was designed. For this, the U/G and T/G mismatches were incorporated into the DNA duplexes of the varying length (Fig. 15). Then, simultaneous and direct measurements of the relative rate of the processing of the T/G and U/G mispairs was performed by the means of the multiple substrate kinetics (Schellenberger *et.al*., 1993). The mismatches were embedded in the sequence context CCXGG, what is the best-cleaved substrate by the Mig.*Mth*II and still quite efficiently processed by the Mig.*Mth*I.

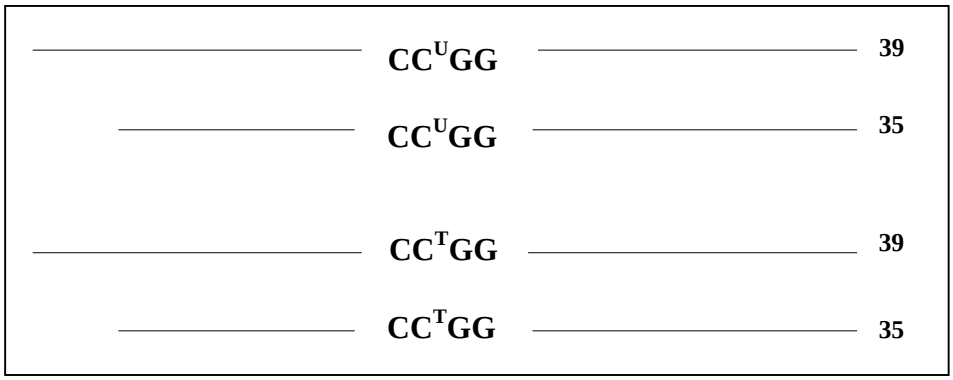


Fig. 15. Schematic representation of the substrates used for the direct U/G, T/G cleavage activity assays.

In the reaction the long substrate with the T/G mismatch was combined with the short U/G containing substrate and *vice versa*. This resulted into two independent assay pairs. As a background control, two substrates only with T/G or U/G mismatches were assayed, to ensure that the length of single stranded overhang does not influence on the cleavage rate. The appropriate measurements were done for the Mig.*Mth*II and Mig.*Mth*I (Fig. 16) under the standard conditions (**4.1.7**). Results are summarised in the Table 11**.**

Enzyme	Cleavage efficiency	U/G:T/U ratio
Mig.*Mth*II	long T/G versus short U/G: k_{rel} = 0.359+/- 0.027	2.8 : 1
	short T/G versus long U/G: k_{rel} = 0.498+/- 0.021	2.0 : 1
Mig.*Mth*I	long T/G versus short U/G: k_{rel} = 0.533+/- 0.045	1.8 : 1
	short T/G versus long U/G: k_{rel} = 0.651+/-0.031	1.5 : 1

Table 11. Direct comparison of U/G and T/G mismatch cleavage by Mig.*Mth*II and Mig.*Mth*I. Relative rate constants are derived from the experiment showed in Fig. 16.

The assay confirmed the rates of the discrimination between U/G and T/G mismatches obtained in the previous experiment. The direct comparison revealed the rate of preference 2 – 2.8 for the U/G over T/G by the Mig.*Mth*II, and about 1.5 times in case of the Mig.*Mth*I enzyme (Table 11). This is in a good agreement with the results obtained from the earlier experiment with the reference sequence. Then, the rate of U/G preference over T/G is 2.1 by the Mig.*Mth*II and 1.2 by the Mig.*Mth*I the CCXGG sequence context (Table 10). As shown in the Fig. 16, the ss overhang on the 5’ end of the substrates has no influence upon the mismatch cleavage activity.

Fondufe, 1999 documented the influence of the origin of the substrates on the activity of Mig glycosylases. It was found that the Mig.*Mth*I is sensitive to the length of the double stranded region of the substrate. Concerning this, all the substrates for the activity assays were designed so, that the respective region would be invariable (31-mer, see **Materials**). The substrates then differed only in the length of the ss overhangs at the 5’ end. However, the influence of the ss overhang upon the substrate cleavage was not tested quantitatively. The scheme of the latter experiment enabled this measurement. No significant influence of the length of the ss overhang on the activity of Mig.*Mth*II and Mig.*Mth*I was detected. This is in a good agreement with the non-specificity of the Mig type glycosylases to the ss DNA.

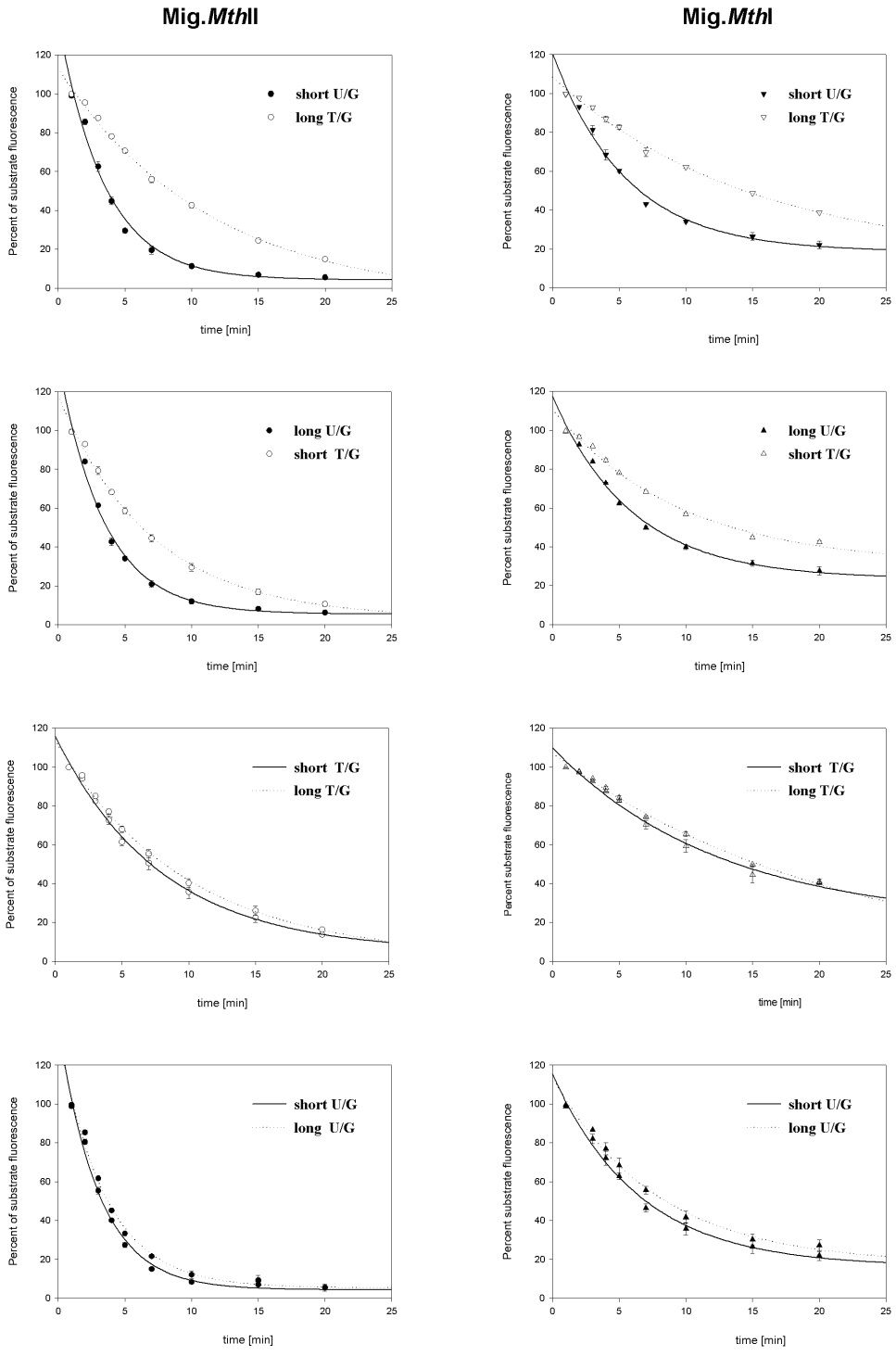


Fig. 16. Direct measurement of U/G and T/G cleavage by the Mig.*Mth*II and Mig.*Mth*I. White colour symbolizes T/G mismatches, black – U/G. Circles show the activity of the Mig.*Mth*II, triangles – Mig.*Mth*I.

4.1.14 Substrate recognition properties of the Mig.*Mth*II and Mig.*Mth*I

Mig glycosylases are characterised to act as the U/G and T/G specific enzymes. However, it is known that Mig.*Mth*I is also active on the mismatches like A/G, T/C, U/C, G/G (Horst and Fritz, 1996). On contrary, PA.Mig enzyme exposes strict substrate specificity by cleaving only U/G, U/oxoG and T/G, T/oxoG mispairs (Yang *et.al.*, 2000). The substrate recognition tests were performed with the Mig.*Mth*II enzyme with the aim to fill in the picture of the plasticity, biological role and requirements for substrate recognition of the MIG enzymes.

The set of substrates were made in the following way: five upper labelled oligos, containing A, C, G, T and U and five lower, non – labelled oligos with the same base variations were hybridised, thus resulting in 25 different mismatches. The mismatches formed were embedded in the CCXGG context, what is the "optimal" sequence for the Mig.*Mth*II (4.1.10). The latter enzyme was incubated with all the heteroduplexes under the standard reaction conditions (4.1.7) and analysed on the A.L.F. sequencer. In parallel, the experiments with the Mig.*Mth*I were performed under the identical conditions. The data were compared with that obtained by the cleavage of various mismatches embedded into GGXCC sequence context (Horst and Fritz, 1996; this work). The summary is presented in the Table 12.

	$GG^{X}CC$	$CC^{X}GG$
Mig.*Mth*II	U/G (suboptimal)	U/G>T/G>>G/G>>U/C,T/C (optimal)
Mig.*Mth*I	U/G>T/G>G/G>>U/C,T/C>A/G (optimal)	U/G>T/G>G/G>>U/C,T/C (suboptimal)

Table 12. Substrate recognition properties of Mig.*Mth*II and Mig.*Mth*I.

The measurements unambiguously indicated that the Mig.*Mth*II is more stringent enzyme than Mig.*Mth*I. Only the U/G and T/G mismatches were cleaved within favourable sequence context. While only U/G is processed in the "bad" sequence context. Besides that, only marginal activity towards G/G was observed (Fig. 17). Very little activity was observed towards cleavage of U/C, U/U, U/T, T/C and T/T mismatches (some of them are not shown).

Mig.*Mth*I is less discriminative enzyme and, as expected, exhibits broader substrate specificity spectrum. Similar to the other Mig glycosylases, U/G and T/G are

the substrates of choice. However, the enzyme is able to accommodate not only pyrimidines, but also purines in the active centre. For example, G/G is processed with about 50% efficiency to that of the U/G cleavage (Fig. 17) and the activity on the A/G mismatches is also detectable (Table 12). Base/base oppositions U/C, U/U or U/T are processed with the marginal activity, similar to the Mig.*Mth*II.

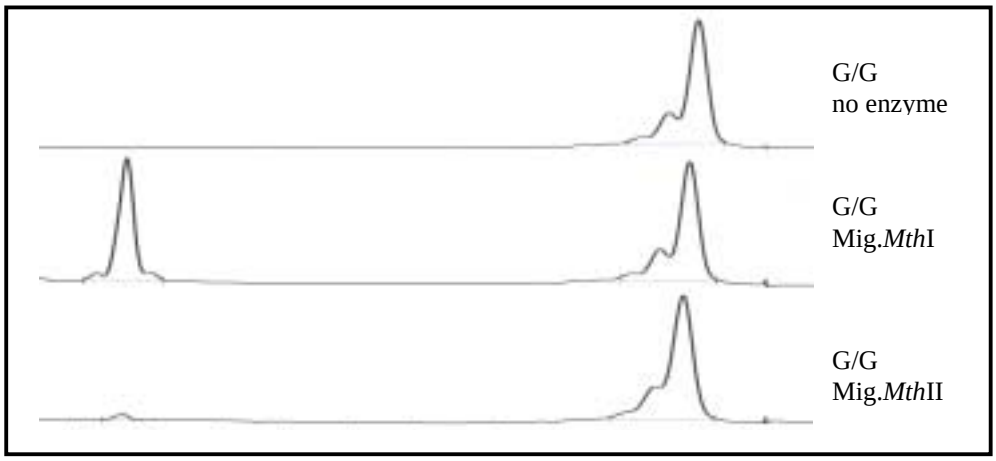


Fig. 17. Differences in cleavage of the G/G mismatch by the Mig.*Mth*II and Mig.*Mth*I.

Remarkably, no U/A cleavage was observed in case of both Mig enzymes in any of the sequence contexts tested.

These data in combination with the requirements for the double stranded DNA substrates (Fig. 12), indicate that Mig glycosylases require specific hydrogen bonds with the orphan base to ensure the normal catalytic activity. Similar observations were for the TDG and MUG enzymes (Barrett *et.al.*, 1998). As the latter glycosylase was crystallized together with the DNA substrate, the specific interactions were determined in great details (Barrett *et.al.*, 1999). The same can be said about the MutY, which crystal structure with the bound adenine was recently published (Guan *et.al.*, 1998). In accordance to that, the amino acids in the minor-groove reading motif α2-α3 were proposed to make specific bonding with the opposite G base, particularly *Gln* 41 residue. Surprisingly, the respective region in the Mig glycosylases is very different, despite that the enzymes require also G opposition. The multiple sequence alignment revealed the high conservation of the *Arg* residue in the respective position (47 a.a. for the Mig.*Mth*II) (Fig. 2), which probably could make the specific hydrogen bonding with the guanine, but not adenine residue. Thus, the pronounced differences in the motives, which are involved in the recognition of the mismatched bases indicates that the enzymes could employ various strategies. However, again in the

absence of the three-dimensional structure of the Mig glycosylases the refined explanation is not possible.

The Mig.*Mth*II and Mig.*Mth*I are more relaxed in the favourable sequence contexts (Table 12). Noticeably, this stringency correlates with the origin of the flipped base, while no effect on the opposition was detected. For example, the Mig.*Mth*I removes A/G in the GGXCC sequence context, but not in the "bad" substrate. Uracil and thymine residues in the opposition with the C are processed marginally in all the substrates tested by both enzymes. Thus, as earlier mentioned, the geometry of the active pocket could change dramatically depending on the specific interactions with the certain DNA sequences. At the same time, that does not influence the bonding with the orphan base.

4.1.15 pH influence upon the activity of Mig glycosylases

The initial activity tests of Mig.*Mth*II revealed that the pH value of the standard reaction buffer with pH 9.0 (room temp.) (Horst, 1996) was not suitable (data not shown). It was found that the Mig.*Mth*II was more active in a buffer with pH 7.5 (room temp., see **Materials**). Therefore, the activity measurements with Mig.*Mth*II described in this work were performed at this pH value.

More accurate evaluation of the dependence of the activity of both Migs upon pH value was tested. The enzymes were separately incubated with several substrates of varying degree of efficiency, containing both T/G and U/G mismatches. Buffers with pH 9.0 and 7.5 were used (though, estimated to be of 8.3 and 6.8 pH values at the standard 50 ^{o}C reaction temp, respectively). TG^{U}CG is presented in Fig. 18 as an example. The respective substrate is only weakly cleaved by Mig.*Mth*II, but efficiently by Mig.*Mth*I.

The experiment revealed that both enzymes act at the lower pH value, preferentially. However, the degree of sensitiveness is very different. The activity of Mig.*Mth*I is not significantly dependent on the reaction pH. At pH 9.0, the reduction of about 1.5-1.7 times in the enzyme activity was observed in comparison to that at pH 7.5. The activity of Mig.*Mth*II, on the other hand, was fully abolished at more alkali conditions (Fig. 18). Similar picture was observed with the "good" substrate CC^{U}GG. The activity of the Mig.*Mth*I dropped down only about twice, while the decrease of the activity of Mig.*Mth*II was more than 5 times (data not shown). Cleavage of T/G mismatches followed the same pattern.

Not only does Mig.*Mth*II exert a high stringency with respect to the sequence context, to the type of mismatch, it also acts within a more narrow pH interval.

Important to note, that determination of the broadness of pH intervals and pH optima for both enzymes is left beyond the scope of this work. The focus was done only within the interval of *in vivo* conditions described for the various strains of *M. thermoautotrophicum* (Jones, *et.al*., 1987). The results obtained served as the guidelines for the optimisation of *in vitro* assays of activities of Mig glycosylases.

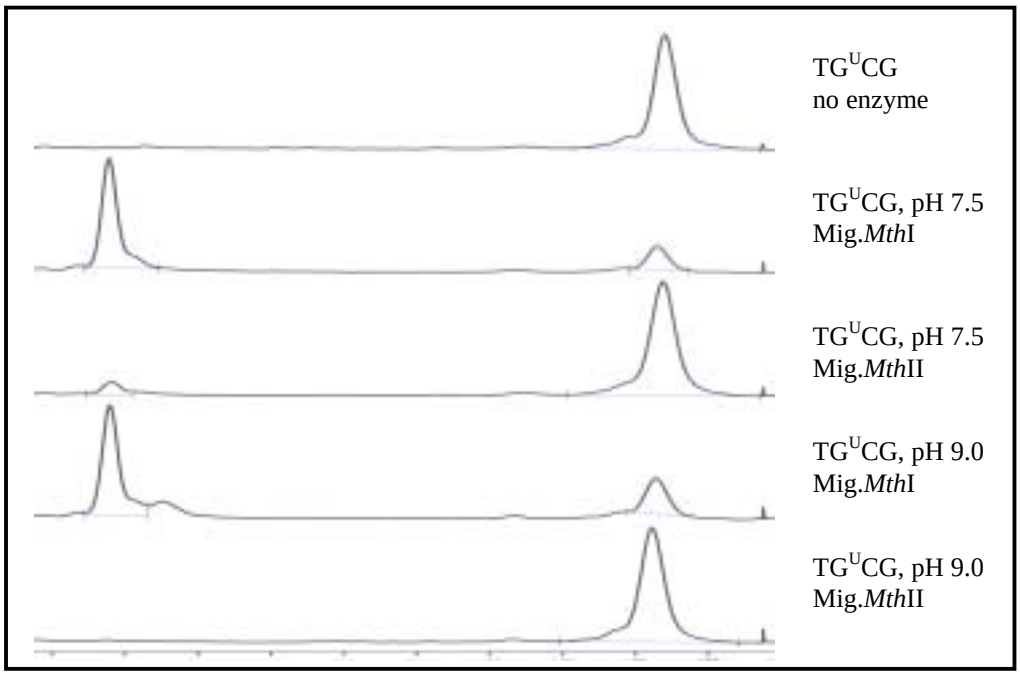


Fig. 18. pH influence upon activity of Mig.*Mth*II and Mig.*Mth*I.

Most likely, the observed differences occur due to the different pI values of the enzymes. Mig.*Mth*I has pI 10.7 (calculated by GCG Sequence Analysis Software Package), then consequently, within the range of pH values tested, the enzyme has a prevalence of positive charges. This possibly facilitates the interaction with the negative DNA molecules. High occurrence of residues of glutamic acid in the composition of Mig.*Mth*II decreases pI till 7.0. Thus, the enzyme has net zero charge in the range of buffers tested, or even obtains negative charge (at pH 9.0), what possibly reduces interaction with the substrate molecules.

4.2 Work on the methyltransferase M.*Mth*H from *M. thermoautotrophicum ΔH*

The quantitative measurements of the specificity of the Mig.*Mth*II revealed that the glycosylase discriminates highly towards the sequences surrounding T/G and U/G mismatches (Table 10). The gene coding for the Mig.*Mth* II is in a physical interaction with *orf* MTH 495, annotated as a putative cytosine - specific methyltransferase (Fig. 4). Consequently, the functional coupling of the activities of two enzymes is expected, similar way as described for Vsr/Dcm pair (Sohail *et.al*., 1993). Then, the detrimental effect of the DNA deamination damage can be reduced by the efficient action of the repair enzyme in the sequence context similar or identical to that of the respective methyltransferase. Thus, the proof of the functional interaction between the DNA methylation and following repair events requires the knowledge about both activities involved.

The detailed enzymatic characterisation of the Mig.*Mth*II glycosylase is presented in this work. Unfortunately, no information is available about the methylation pattern of the genome of *M.thermoautotrophicum* ΔH (Smith *et.al*., 1996; Roberts *et.al*., 1998). The same can be said about the activity of the putative methyltransferase coded by the *orf* 495.

For further work, the product of the *orf* 495 was named M.*Mth*H methyltransferase in accordance to the standard nomenclature. The same as present in the REBASE^{TM} database.

4.2.1 Search of the homologs of the M.*Mth*H methyltransferase

First a search for homologs of M.*Mth*H methyltransferase was performed.

In general, three types of the site-specific methyltransferases have been described: cytosine-specific on either the C5-position or the N4-position, and adenine-specific on the N6-position. The enzymes can be easily differentiated according to the composition and distribution of their conserved motives (Malone *et.al*., 1995). Particularly related are the 5-cytosine-specific enzymes (Kumar *et.al*., 1994). Unfortunately, the part of the enzyme, which recognises the DNA sequence, is highly variable. Because of this, *a priori* prediction of the methylation pattern is almost impossible.

However, accumulation of more experimental data about different methyltransferases enabled a prediction of the unknown activities of some proteins. To achieve this, the sequence of the variable region should have no less than 50-60% of the identity to that of the characterised enzyme. This approach was used for the prediction of the activity of the M.*Mth*TI located on the plasmid of *M.thermoautotrophicum* THF with the following experimental confirmation (Nölling and de Vos, 1992). The enzyme has very strong similarity to the other GGCC pattern specific methyltransferases.

Unfortunately, the target recognition domain of the M.*Mth*H is too divergent from the subfamily of GGCC specific methyltransferases (data not shown). Therefore, BLAST search for the closest homologs of the M.*Mth*H was performed among the characterised enzymes. The search was done in two ways. The sequence of the full-length protein or only the sequence string of the variable region was used as a query. In both approaches, the hit with the highest score was to the methyltransferase from bacteria *Agrobacterium gelatinovorum* M.*Age*I. The search revealed that the variable region of the M.*Mth*H is unique. The output of the search with the sequence of the variable region resulted in the only hit, which was not extended even by several Ψ-BLAST iterations:

```
M.MthH   97  PPELRTHRNTRSFLDRYKVVAGDLPYSHTIVAHISKDGHYYIHP 140
             P ++    +++ F ++Y+ +  D P SHT+VAH+++D   ++HP
M.AgeI  332  PAKVMEFLDSQGFANKYRRLRWDAP-SHTVVAHMARDCSDFVHP 374
```

The string demonstrated here spans only 43 a.a. out of 140 residues comprising the whole variable region. 31% of the identity and 65% of the similarity obtained are not sufficient for the reliable prediction of the methylation pattern. On the other hand, M.*Age*I recognition sequence is ACCGGT. This is in a good agreement with the kinetics data of Mig.*Mth*II with CCGG sequence serving as the most efficient substrate.

Unfortunately, the latter information can be used only as a guideline for the further work on the M.*Mth*H. The cloning of the respective gene and experimental identification of the recognition sequence is required.

4.2.2 Cloning the gene of methyltransferase M.*Mth*H

The immediate isolation of the intact gene MTH 495 by the PCR reaction from cell suspension of *M. thermoautotrophicum* (see **4.1.4**) was not successful. No PCR

product was obtained despite the numerous optimisation steps. Therefore, a different cloning strategy was applied (Fig. 19).

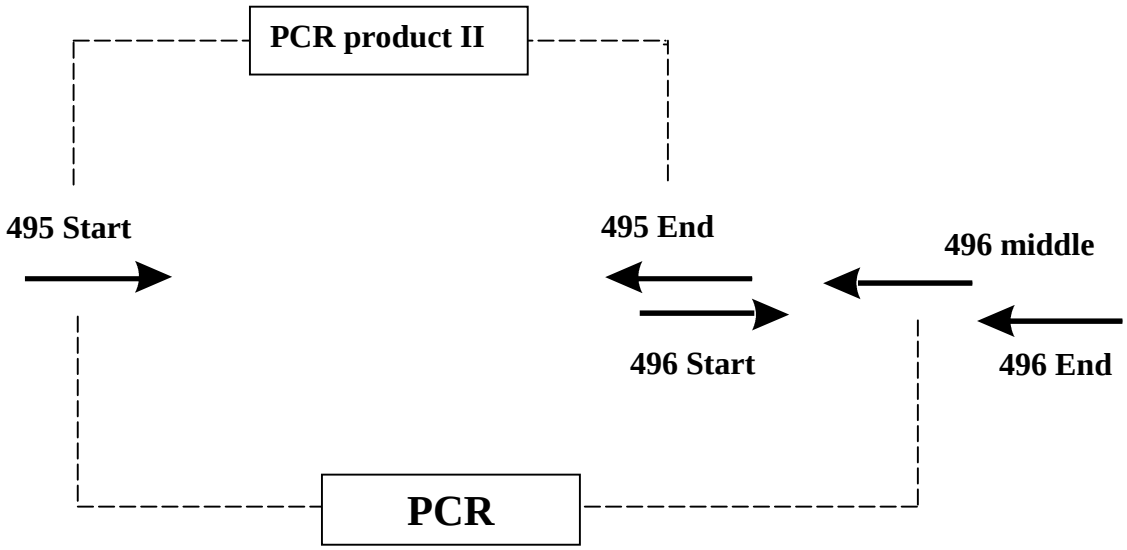


Fig. 19. Scheme for cloning the gene of M.*Mth*H from *M.thermoautotrophicum*. For more details see text.

Fortunately, the primer seqMTHmid (located within the *mig*2 gene and here indicated as the arrow 496 middle) in combination with Ncometilaze primer (here shown as an arrow 495 Start) resulted in the fragment (PCR product I), which was used for the further manipulation. Surprisingly, no second targeted PCR product was obtained from this DNA. In addition, the first PCR reaction was very inefficient and the product was hardly visible on the agarose gel (data not shown). Thus, for the isolation of the right-length MTH 495 gene, the PCR product I was subjected to the blunt - end cloning into the pCR - Blunt II - Topo vector (Invitrogen). The resulting clones were screened by the PCR reaction and the plasmid DNA purified. The sequencing of the ends of the insert confirmed the correctness of the construct. Finally, using the appropriate plasmid DNA as a template, PCR product II (Fig. 19) was obtained and cloned into the expression vector pET-21d. The correctness of the MTH495 gene (1239 bp.) was confirmed by the sequencing with the flanking (T7 and T7 terminator) and the walking primers (seqMTHconf and Mthmetmidb).

4.2.3 Expression studies of M.*Mth*H methyltransferase

Before the activity assays and protein purification procedure, the expression test of the M.*Mth*H was performed. For this, the BL21(DE3)pLysS cells with the respective plasmid were grown in 2 x YT media with the appropriate antibiotics (see **Materials**) until OD_{600} reached 0.6 - 0.7 at 37^0C temp. Induction of the protein synthesis was started with 1mM IPTG and the samples were collected every hour. Cells were collected by centrifugation, resuspended in the 20 mM HEPES, pH 7,6 buffer and disrupted by the sonication. The soluble and non – soluble fractions were separated by the centrifugation and loaded on the 12% SDS protein gel. However, no visible appearance of the band with the expected size of 49.4 kDa was observed in the soluble fractions (data not shown). Therefore, the Western-blot was performed and the protein was visualised with the penta – His tag antibodies and the chromogenic substrate BCIP/NBT (Fig. 20).

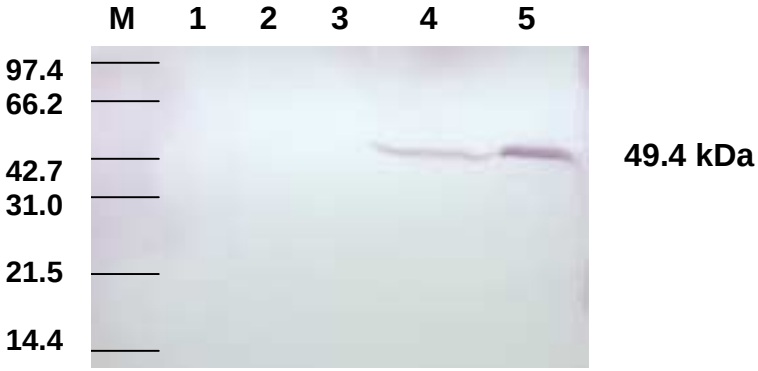


Fig. 20. Western blot of the M.*Mth*H. M – Mid – range protein molecular marker (Promega), 1 – supernatant before induction, 2 – supernatant after 3h induction, 3 – pellet before induction, 4 and 5 – pellet after 3h induction in two independent experiments. The staining was done with AP – conjugated secondary antibody and NBT/BCIP.

The experiment revealed that the M.*Mth*H methyltransferase is expressed by the cells, but present only in the debris. This is, however, a common than exceptional case for the recombinant methyltransferases (Kumar *et.al*., 1992). Frequently, these enzymes are not able to fold properly in the cytoplasma of *E. coli* and have special requirements to the appropriate salts and their concentration, pH values. Alternatively, the protein can be trapped into inclusion bodies. To check this, cells were observed under the light microscope after induction of the protein synthesis. No inclusion

bodies could be detected. Therefore, the next experiment was designed to extract and solubilise the M.*Mth*H from the cell pellet under the mild conditions.

4.2.4 Experiment to extract M.*Mth*H methyltransferase from the cell pellet

The experiment was designed to solubilise the precipitated M.*Mth*H methyltransferase under different pH, salt and buffer conditions. Similar approach was described by Kumar *et.al*., 1992 and used for the production of the M.*Hha*I enzyme.

The experiment was done the following way: the cells were incubated at 37 ^{o}C until OD_{600} = 0.7, then induction of the protein synthesis was started with 1 mM IPTG. After 3 hours the cells were collected by centrifugation and lysed. Soluble and non-soluble fractions were separated. The pellet was repeatedly solubilised in the buffer, containing various concentration of NaCl (200 mM and 400 mM). The suspension was separated into the soluble and non-soluble fractions by centrifugation. The fractions were applied on the 12% SDS-protein gel. In parallel, three different buffers systems were used: 20 mM HEPES, pH 7.6, 50 mM Tris-HCl, pH 8.0 and 10 mM phosphate, pH 7.5. Results are shown in the Fig. 21.

Unfortunately, the *M.Mth*H remained in the pellet, despite the type of buffer and NaCl concentration used. The reasons to this are hard to explain. The correctness of the construct pET-21d/M.*Mth*H was checked by sequencing. Full coverage of the methyltransferase gene was achieved with no mistakes in the nucleotide sequence. The reaction with the anti-His tag antibodies confirmed (Fig. 20) that the insert stands “in frame”, and the protein is translated.

Possibly, the protein can be naturally inactive, due to the observation that a lot of genes coding for restriction-modification systems are non-functional (Roberts, 1998). However, the comparison of the sequence of the M.*Mth*H with the other functionally characterized methyltransferase did not reveal any unusual deviation (data not shown). The protein contains all ten conserved motives in the correct orientation (Kumar *et.al*., 1994) and the respective residues involved in the catalysis, like the thiol-group donor Cys 122 (Hanck *et.al*., 1993). Therefore it is very unlikely that M.*Mth*H would be functionally inactive. The precipitation of the enzyme probably occurs due to non-favourable conditions, rather than to “programmed” inactivation.

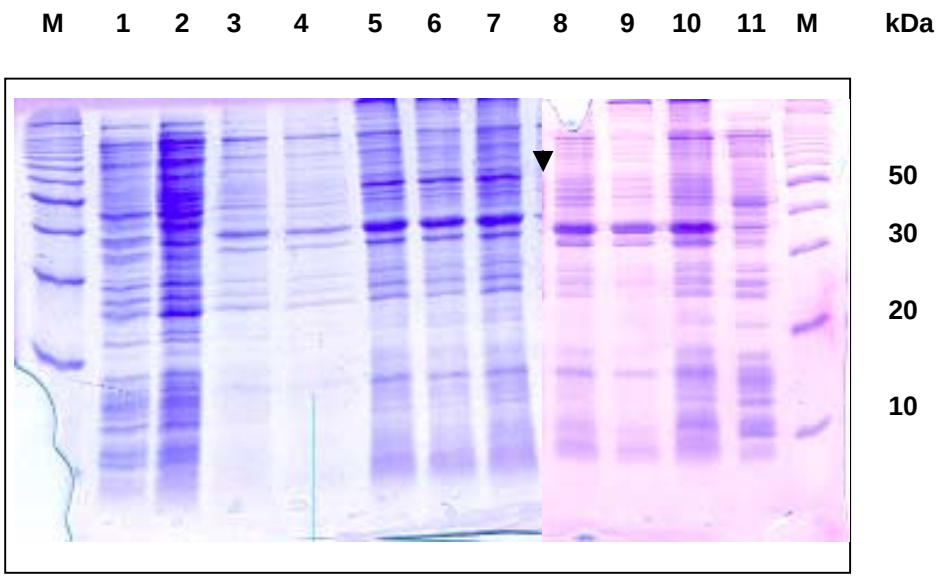


Fig. 21. Experiment for the extraction of *M.Mth*H methyltransferase from the cell pellet. **Lane 1:** cells before induction, **lane 2:** supernatant after wash without NaCl, **lane 3:** supernatant after wash with 200 mM NaCl conc., **lane 4:** supernatant after wash with 400 mM NaCl conc., **lanes 5, 6, 7:** pellet after washing with 0, 200, 400 mM NaCl conc., respectively, **lanes 8, 9, 10:** pellet of the cell culture carrying the pET vector without insert (control) after wash with 0, 200, 400 mM NaCl conc., respectively, **lane 11:** control cell culture before induction. **M:** 10 kDa protein size marker, GIBCO. Experiment shown was performed in 20 mM HEPES, pH 7.6 buffer. Analogous pictures were obtained also in other buffer systems (see text).

4.2.5 *In vivo* activity test of the M.*Mth*H methyltransferase

Despite that vast majority of M.*Mth*H is insoluble in *E. coli* (Fig. 21), likely that soluble fraction of is also present. Then, even the negligible amounts of the protein should be sufficient to methylate DNA in the detectable range. Therefore, *in vivo* methylation test was designed.

For this, the standard expression system (BL21(DE3)pLysS/pET-21d/M.*Mth*H) was used, since BL21(DE3)pLysS is a dcm$^-$ strain (Novagen). Although dam methylase is present, the interference between the adenine 6 and cytosine 5 methylation patterns is also unlikely. The growth of the cells and induction of the synthesis of the M.*Mth*H was done under standard conditions. If the expressed enzyme is active, it should be able to transfer the methyl-group from the intracellular S-adenosyl-L-methionine (Gold *et.al*., 1964) on to the genomic and plasmid DNA. The latter was purified by the Qiagen kit and subjected to the restriction digest

Restriction endonuclease	Recognition sequence	Sensivity to C-5 methylation
*Bsa*WI	**W*CCGGW**	WC ^{m5}CGGW
*Cfr*10I	**R*CCGGY**	RC ^{m5}CGGY, R ^{m5}CCGGY
*Eco*RII	***CCWGG**	^{m5}CCWGG, C^{m5}CWGG
*Sac*II	**CCGC*GG**	^{m5}CCGCGG, C ^{m5}CGCGG, CCG ^{m5}CGG
*Nco*I	**C*CATGG**	^{m5}CCATGG
*Eag*I	**C*GGCCG**	^{m5}CGGCCG, CGGC ^{m5}CG
*Apa*I	**GGGCC*C**	GGGCC ^{m5}C
*Ava*II	**G*GWCC**	GGW ^{m5}CC, GGWC ^{m5}C
*Eco*0109I	**RG*GNCCY**	RGGNC ^{m5}CY
*Ehe*I	**GGC*GCC**	GG ^{m5}CGCC, GGCG ^{m5}CC
*Bam*HI	**G*GATCC**	GGAT ^{m5}CC
*Pvu*I	**CGAT*CG**	CGAT ^{m5}CG, CG methylation sensitive
Bgl II	**A*GATCT**	AGAT ^{m5}CT
*Bss*HII	**G*CGCGC**	CG methylation sensitive
*Mlu*I	**A*CGCGT**	A ^{m5}CGCGT, ACG ^{m5}CGT
*Hae*II	**RGCGC*Y**	RG ^{m5}CGCY, RGCG ^{m5}CY
*Eco*47III	**AGC*GCT**	AG ^{m5}CGCT
*Fsp*I	**TGC*GCA**	TG ^{m5}CGCA
*Xho*I	**C*TCGAG**	^{m5}CTCGAG
*Pst*I	**CTGCA*G**	^{m5}CTGCAG, CTG ^{m5}CAG
*Ecl*136II	**GAG*CTC**	GAGCT ^{m5}C
*Pvu*II	**CAG*CTG**	CAG ^{m5}CTG
*Bsr*BI	**CCG*CTC**	^{m5}CGCTC, C ^{m5}CGCT ^{m5}C
*Bst*1107I	**GTA*TAC**	GTATA ^{m5}C
*Hpa*I	**GTT*AAC**	GTTAA ^{m5}C
*Psp*1406I	**AA*CGTT**	AA ^{m5}CGTT
*Nsi*I	**ATGCA*T**	ATG ^{m5}CAT
*Bsa*AI	**YAC*GTR**	YA ^{m5}CGTR

Table 13. Restriction enzymes used for analysis of M.*Mth*H activity. Grey boxes indicate enzymes that were cleaving only partially. W = A or T, R = G or A, Y = C or T, N = G, A, T, C.

analysis. The activity of the M.*Mth*H was determined by the non-ability of the restriction endonuclease to incise the DNA. 30 different cytosine 5 methylation sensitive restriction endonucleases were used. Results are presented in the Table 13.

The digest analysis of the plasmid revealed four enzymes, which cut the DNA only partially. Two of them – *Pvu*I and *Bgl*II – shear the recognition sequence "GATC". Unfortunately, one more "GATC" pattern recognizer, *Bam*HI, was not inhibited. The sequence is the target for the dam methylase (Brooks *et.al.*, 1983), which is present in the strain BL21 (DE3) pLysS. However, the restriction endonucleases used are sensitive only to the C5 methylation and as stated before the interference between dcm and dam is unlikely.

Another two not-fully cutting enzymes, *Apa*I and *Bss*HII, have quite different recognition sequences with the clear prevalence of the GC bases. Unfortunately, no consensus sequence can be derived from the data of this experiment. Surprisingly, none of the enzymes with the specificity to the "CCGG" pattern, what was experimentally determined to be the best cleaved sequence by the Mig.*Mth*II, was retarded in the activity.

The restriction sites of these enzymes are found once or twice on the plasmid and the cleavage reaction could be hindered by a secondary structure. The following experiment was performed to clarify this. The reactions with four not – fully cutting enzymes were repeated with the prolonged incubation time up to 4 h. The analogous plasmid from the non – induced culture was used as a negative control. The results are presented in Fig. 22. Also with the prolonged incubation times there was no influence on the activity of the restriction endonucleases used. On the contrary, the control plasmid was fully digested under the same conditions. Therefore, the most likely, the observation is that there is some of methylation activity in the induced culture. The fact that DNA restriction was not inhibited fully can be explained by only partial activity of the thermostable methyltransferase M.*Mth*H at 37 ^{o}C temperature. Similarly, only partial digest of the methylated DNA was reported during the enzymatic characterization of M.*Mth*TI (Nölling and Vos, 1992).

The observed possible activity of the M.*Mth*H is in a reasonable agreement with the patterns of methylation of other *Methanobacterium* species. Interestingly, to note, that *M. thermoautotrophicum* ΔH could have reversed methylation patterns in comparison to other characterized representatives of the group. "GATC" is the direct reversal of the "CTAG" detected in *M. thermoautotrophicum* FTF and Z-245. Majority of the other strains: *M. thermoautotrophicum* RC, THF, BR10 carry "GGCC" pattern. The only different pattern "CGCG" was described in the *M. vannuli* (data taken from REBASE database, (Roberts and Macelis, 2001)).

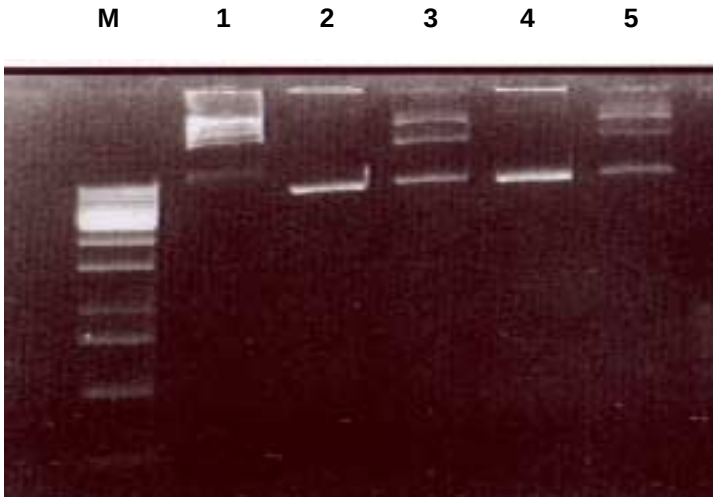


Fig. 22. Examples of the repeated restriction digest with the partial cutters and prolonged incubation times. **M**: GeneRuler 1kb DNA Ladder (MBI Fermentas), **1**: control – non – digested pET-21d/M.*Mth*H, **2**: control after 1h incubation with *Bss*HII, **3**: possibly methylated plasmid after 1h incubation with *Bss*HII, **4**: control plasmid after 4h incubation with *Bss*HII, **5**: possibly methylated plasmid after 4h incubation with *Bss*HII.

Unfortunately, the experiment was not sufficient for the unambiguous determination of the methylation sequence. Alternatively, the multi-functionality of the M.*Mth*H could be proposed. These enzymes usually contain several clearly spaced target recognition regions (TDR), which are responsible for the interaction with the different sequences (Behrens *et.al.*, 1987). The gene structure of M.*MthH* indicates the presence of only one TDR, therefore, the protein is, most likely, single-functional.

Possibly, more specific answers could be obtained by the *in vitro* assay at the appropriate temperature of 50-65 ^{O}C. This can be done with the extract of the induced cells (Nölling and Vos, 1992). Alternatively, the use of the different expression system can be required to obtain soluble protein.

Despite the problems encountered, further work on this methyltransferase would be very useful.

4.3 Analysis of the gene loci of the microorganisms containing Mig-type glycosylases

4.3.1 Distribution of the Mig glycosylases

The U/G, T/G specific Mig glycosylases have only limited distribution and are present only in 7 organisms among 273 sequenced ones (WIT/restricted database). All representatives found to date are microbial enzymes (Fig. 2). In most cases, they are spread among archaea with the only exception: green, non-sulfur bacteria *Chloroflexus aurantiacus*. Mig glycosylases are prevailing among thermophilic organisms: *Aeropyrum pernix* and *Pyrobaculum aerophilum* are extrathermophiles with growth temperatures above 95 ^{o}C, *Methanobacterium* strains are moderate thermophiles with the optimal growth temperatures at 65 ^{o}C. *Methanosarcina mazei* is the only mesophilic organism. Among the hosts is *Ferroplasma acidarmanus*, which was described as the most acidophilic organism known to date (microbiological data taken from Genome On Line Database GOLD^{TM} , Bernal, *et.al*., 2001).

4.3.2 Surrounding of the genes of Mig glycosylases

Genetic loci of the Mig glycosylases in all microorganisms were compared with the aim to get a detailed insight into the conservation of the closest environment of the enzyme and to determine the biological role, if possible **(**Fig. 24**)**.

Recently, the Mig-type glycosylases were isolated only from *M.thermoautotrophicum* strains (Horst and Fritz, 1996; this work). Consequently, Mig.*Mth*I and Mig.*Mth*II were the closest and the only homologs for some time. With the accumulation of the genomic data, more candidates were found (Yang *at.el*., 2000). Comparison of their sequences revealed the surprise – the closest homolog of Mig.*Mth*II is present in the genome of another methanogen - *M.mazei* (unpublished). Both proteins share about 50% of the identity and 63% of the similarity on the level of a.a. (GCG Sequence Analysis Software Package). The similarity to the Mig.*Mth*I is somewhat less with 32% identical and 43% similar residues. The high degree of conservation is also observed in the closest surrounding of the Migs: the upstream genes, transcribed into the same direction also are 5-cytosine specific methyltransferases. Not only are the Mig.*Mth*II and MM.Mig enzymes closest homologs, but also the sequences of the variable regions of both methyltransferases are highly conserved. Due to the similarity between their protein sequences, it would

be reasonable to expect that the methylation pattern could be well conserved between both methanogenic organisms (S.Klimasauskas and S.Serva, personal communication).

The two other members of this group of Mig glycosylases are from the extraacidophile *F.acidarmanus* (FA.Mig) and thermostable bacteria *C.aurianticus* (CAU.Mig) (unpublished). FA.Mig shares about 50% of the identity and more than 60% of similarity to Mig.*Mth*II. The methyltransferase is also highly conserved (48% of the identity and 63% of the similarity) with the probability of having the same methylation pattern as the methanogenic proteins. In case of CAU.Mig, only the glycosylase possesses high degree of conservation, including the possible variable region (Fig.2). Thus, the similar substrate specificity spectrum to that of the Mig.*Mth*II is expected. Interestingly, there are several restriction endonucleases described from various strains of *Ch.aurianticus.* Majority of them bear specificity to the "CCGG" patterns (data taken from the REBASE). Thus, despite that the upstream located gene of the methyltransferase has a completely different sequence of the variable region, it could be also specific to the above sequence pattern.

Surprisingly well - preserved is a mutual localisation of the methyltransferase and the Mig glycosylase in all above listed organisms. The stop codon of the methyltransferase serves as start codon for the glycosylase. All the glycosylases, except the Mig.*Mth*II utilize "*ATG*" as a start codon. This confirms the correctness of the choice of the start codon for the Mig.*Mth*II as earlier described. Despite that it is a "*CTG*" start codon (Fig. 5), it also overlaps with the stop codon of the respective methyltransferase gene (*orf* 495) in the same way as the "*ATG*" codons of other glycosylases.

The only difference among the loci is brought by the presence/or absence of the putative restriction endonuclease. Only *M. thermoautotrophicum* contains the putative *orf* 494 for the restriction endonuclease (located upstreams the M.*Mth*H gene). In *M. mazei* this corresponding *orf* is clearly missing, while for *F. acidarmanus* and *C. aurianticus* the sequencing data of the respective contigs are still not available.

The high sequence similarity among this group of Mig glycosylases enables to superimpose experimentally tested stringency of the substrate specificity of the Mig.*Mth*II to the other members of the sub-group. In addition, conserved genomic localisation of the glycosylase genes to the physical proximity to the respective methyltransferases leaves no doubt about the main function of these enzymes. Obviously, this is to repair the T/G mismatches resulting from the deamination of the methylated cytosine residues. Cleavage of the U/G mismatches, most probably, serves as an additional backup activity.

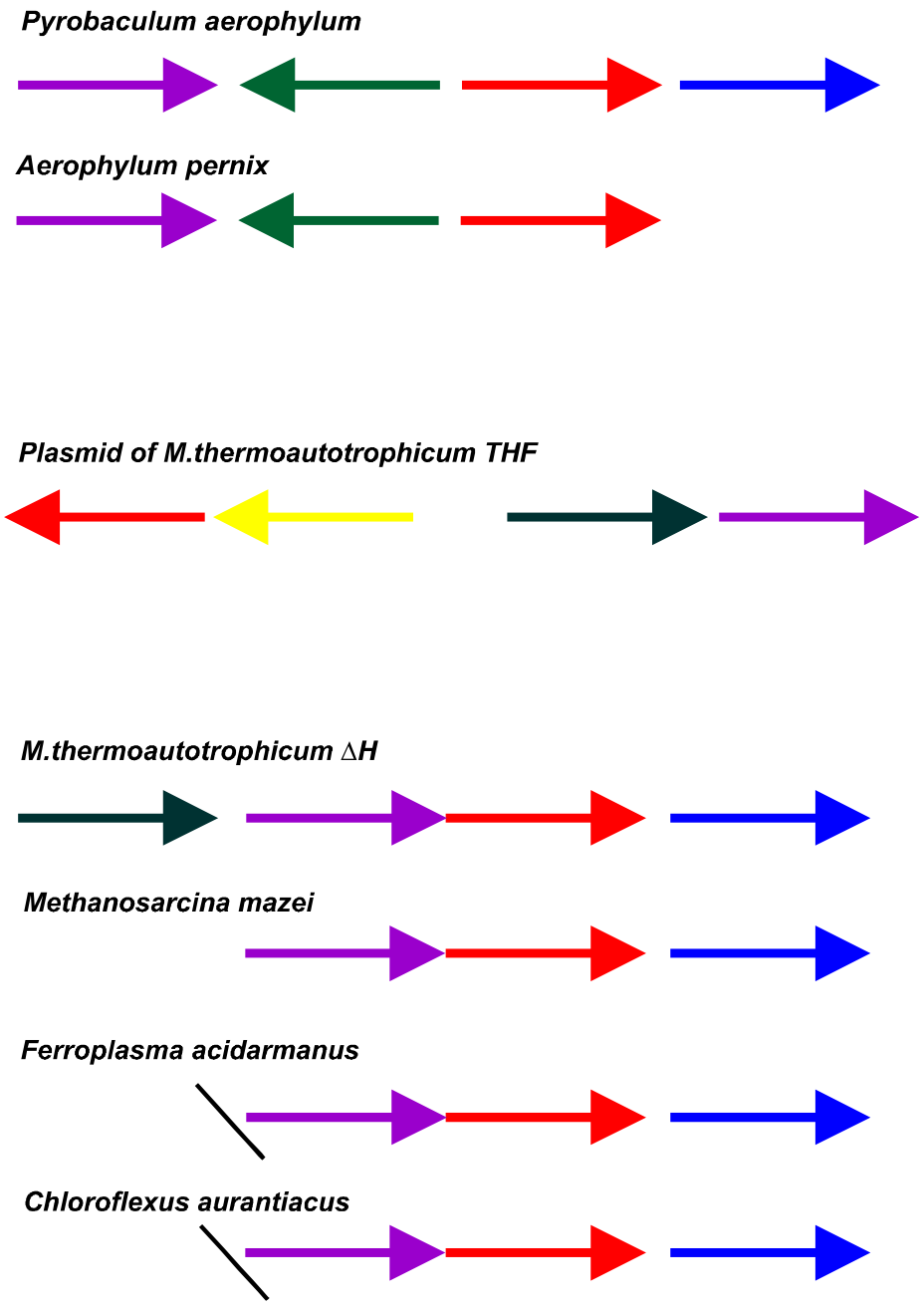


Fig. 23. Genomic loci of the microorganisms containing Mig glycosylases. The arrows indicates the direction of transcription. Colours symbolises genes located in the region of interest: red - is Mig glycosylase, violet – methyltrasferase, black– restriction endonuclease, green – hypothetical protein, most likely restriction endonuclease, yellow – *orf* 9, protein with unknown function, blue – protein, containing DEAD box, annotated as a putative helicase. The black lines mean the shortage of the sequencing data. The data are taken from WIT/restricted database (Integrated Genomics).

Loci of the another two organisms: extrathermophiles *Aeropyrum pernix* (Kawarabayasi *et.al*., 1999) and *Pyrobaculum aerophilum* (unpublished) are also similar. Their glycosylases (APE.Mig (APE 0875) and PA.Mig) share especially high degree of the conservation: about 60% of identity and more than 70% of similarity on the level of protein sequences. Less conservation is observed to the Mig.*Mth*II: about 35% of dentity and about 50% of homology. Highly similar are the methyltransferases (APE 0872 in *A. pernyx*) of these two organisms, also their mutual localisation. The *orf*s (APE 0874 in *A. pernyx*) located between the methyltransferases and glycosylases are highly conserved and possibly could be the restriction endonucleases. Likely, that these genes form the functional unit – operon. However, according the present day definition (Lathe III *et.al*., 2000), the genes belonging to the operon should be located in the same transcriptional orientation and be separated by less than 250 bp. Only the genes of Mig's and methyltransferases are transcribed into the same direction. The *orf*s between them have the opposite orientation. Despite that, the genes of Mig's and methyltransferases are separated by the distance more than 1000 b.p., they can still possibly act in the concerted manner. However, the stringency of the possible interaction is not known.

According to the genomic data, the repair of the T/G mismatches can be attributed as the primary role of the PA.Mig and APE.Mig glycosylases. More general uracil repair pathway, likely, is performed by the uracil glycosylases (TMUDG-type), present in the genomes of *A. pernix* and *P. aerophilum* (unpublished*).*

The third member of the triad is the plasmid-encoded Mig.*Mth*I (Horst and Fritz, 1996). It shares about the same level of identity (32%) and similarity (about 45%) to all other Mig glycosylases. The localisation and transcriptional direction (Nölling *et.al*., 1992) does not indicate the close relationships with the methylation events in the cell. Experimental data also confirmed little specificity to the sequence context of this enzyme (Fondufe, 1999; this work). Thus, it seems as if Mig.*Mth*I glycosylase could serve as a more general DNA repair enzyme. On the other hand, it is unlikely that the plasmid-encoded enzyme performs the function of the general uracil glycosylase. Possibly, it could act as a complementing activity to the chromosomal version of the Mig glycosylase **(4.1.11)**. Another possible function could be as a carrier of the gene of the glycosylase between the different taxonomic groups. Recombination and spontaneous mutagenesis could promote the formation of the chromosome-located enzymes highly adapted to the local conditions.

Finally, in difference to chromosomal loci, the genes of the methyltransferase and the glycosylase on the plasmid of *M. thermoautotrophicum* are interspersed not

only by the restriction endonuclease, but also by *orf* 9 (Nölling *et.al*., 1992) – protein with unknown function (Horst, 1996; Fondufe, 1999). The latter has species-specific distribution. The only homologs of *orf* 9 (*orf* 308 and *orf* 471) are present only in the genome of *M. thermoautotrophicum* ΔH. However, they are located far away from the genes coding for the DNA repair proteins. Thus, it is likely that the close proximity to the Mig.*Mth*I on the plasmid pFV1 (Nölling *et.al*., 1992) could be accidental.

Thus, analysis of the genomic surrounding of the Mig glycosylases in combination with the experimental data revealed that the enzymes are involved in the repair of the products of the methyl-cytosin deamination. The presence of the "true" uracil glycosylases in the genomes of *M. mazei*, *A. pernix* and *P. aerophilum* confirms this suggestion (unpublished). However, some of the organisms are still missing a candidate for the general uracil repair pathway (Eisen and Hanawalt, 1999). For example, no homolog of known uracil glycosylase has been found in the genome of *M. thermoautotrophicum* ΔH. Alternatively, there are the organisms (*Pyrococcus* strains (unpublished)), which contain the genes of the C-5 cytosine specific methyltransferases, but no system for the repair of the resulting T/G mismatches. Thus, not all the elements of the repair of the DNA deamination damages are deduced and there is a lot of space for further research.

4.3.3 Why do the genes of methyltransferases and glycosylases overlap?

The analysis of the loci of the Mig glycosylases raised few intriguing possibilities. Roberts, 1995 had noticed that most of the DNA binding proteins exploit the base flipping mechanism. This enables separating target recognition and base modification (methyltransferase) or repair (glycosylase, exonuclease III) in an easy and comfortable way. In most cases, the DNA repair proteins are small, simple molecules. On the contrary, the DNA methyltransferases are clear examples of the fusion of the several activity modules in one complex protein (Cheng and Blumenthal, 1999). Based on that, the evolution of the DNA methyltransferases from the primitive repair enzymes was proposed (Roberts, 1995).

Despite the absence of any significant sequence similarity between domains of DNA methyltransferases and glycosylases, the stringency in the sequence recognition could serve as a connecting feature. Mig glycosylases therefore could be the missing link between the highly specific methyltransferases and practically unspecific helix-hairpin-helix glycosylases (MutY, for example). The latter being an earlier

development with the later gain of more specialised adaptations. This is favoured by the recent finding, that all the organisms sequenced till date contain the genes of glycosylases of the respective type (Aravind *et.al*., 1999), which, consequently, appears to be the most conserved DNA repair protein. In addition, the connection between the glycosylase and methyltransferase can be hypothesised based on the fact that the latter enzyme exposes high activity also upon the U/G mismatches (Klimasauskas and Roberts, 1995).

However, no reasonable explanation exists, why some of the Mig glycosylases possess high stringency towards the sequence context. The enzyme could perfectly perform its protective function against T/G damages without restrictions to the particular substrates. This was described for the functional homolog of Mig, Vsr endonuclease. The enzyme discriminates against DNA substrates, however, the rate of the discrimination is only about 10-20 times and, in principle, all DNA substrates were processed (Gläsner *et.al*., 1995). Recent crystal structure of the Vsr (Tsutakawa *et.al*., 1999) revealed that the enzyme bears strong similarity to the class II restriction endonucleases. Consequently, the evolution of the Vsr/Dcm coupling from the restriction/modification activities can be proposed. Then, knowing the stringency of the restriction endonucleases to their substrates, the relaxed mode of the Vsr is rather unusual.

Unfortunately, no such predisposition exists for the Mig enzymes. The close and highly conserved localisation of the Mig.*Mth*II subtype of glycosylases and the genes of the methyltransferases, more likely, is the result, but not the reason of the functional coupling. At present, both enzymes, apparently, are transcribed as separate units. However, their close proximity could be a reminiscent of their previous co-localisation as a single polypeptide. The gene fission events were described to occur quite frequently in the thermophilic organisms, particularly by the introduction of the frameshift (Snel *et.al*., 2000). The dissection of complex protein into the subunits could be advantageous in the high temperatures (Jaenicke and Boehm, 1998), nevertheless, not loosing the functional coupling. The gain of the distance between the two genes and/or appearance of the restriction endonuclease (*P. aerophilum* or *M. thermoautotrophicum* THF) could represent the course of the following evolutionary events and the functional divergence.

Some other scenarios could be suggested to explain the conserved overlap between glycosylase and methyltransferase genes. Namely, the methyltransferase –

glycosylase fusion protein can be produced, despite the presence of the frameshift. The latter could possibly be avoided during the translation process (Nelson and Cox, 2000). Remarkably, all the overlapping methyltransferases carry "TGA" stop codon. The latter is known to be most frequently involved in the regulated frameshifting and serves as a signal for the production of the extended proteins (Tate *et.al*., 1996). In addition, the frameshift of one base is characteristic for the regulated bypass (Baranov *et.al*., 2001). Alternatively, posttranscriptional editing of the mRNA (Brennicke *et.al*., 1999) could be also proposed.

The presence of several chimera proteins indicates that this strategy is used for conferring additional specificity to the helix-hairpin-helix glycosylases. Known examples are bacterial MutY and AlkA glycosylases. In case of MutY, the core is fused with the 8-oxo-dGTPase (Noll *et.al*., 1999), while for AlkA, the fusion occurred with the ADA protein, specific for the recognition of the methylating agents (Bond and Wallace, 1999). Recent work of Hendrich *et.al*., 1999, described the fusion of the helix-hairpin-helix glycosylase with the methyl-CpG-binding domain in humans.

4.3.4 DNA and RNA helicases – new players in the base excision repair?

The blue arrows in the Fig. 23 indicate *orf*s, annotated as class II ATP-dependant RNA and DNA helicases containing DEAD box motives (Bird *et.al*., 1998). *Orf*s in *M. thermoautotrophicum* (MTH 497), *M. mazei*, *F. acidarmanus* and *C. aurianticus* shows high conservation in a.a. sequences (about 40% of the identity and more than 50% of the similarity). The *orf* in *P. aerophylum* has more divergent composition, but still can be aligned together with the other putative helicases through the whole length of the protein. The close localisation and the same transcriptional orientation places the genes of the putative helicases in the same operon together with the DNA repair proteins.

Unfortunately, the actual activity of these putative helicases cannot be elucidated from the sequence similarity data. In general, the group of DEAD boxes containing proteins are involved in various DNA and RNA metabolic processes, like protein-DNA interactions, transcription regulation and translation (de la Cruz *et.al*., 1999). Due to the difficulties of monitoring of their *in vivo* activities, there is very huge lack of the biochemical data. Consequently, the role of the putative helicase in the action of the Mig glycosylases can be just speculated.

If the helicase works on the DNA substrates, it can be directly involved in the base excision repair pathway. Despite that this mechanism is not known to involve any DNA unwinding activities (Parikh *et.al*., 1999), the latter possibility can not be rejected. Especially since the actual row of events in procaryotes, especially in archeae is not known. If the repair of damaged DNA strand goes according to the long-patch mechanism, then the helicase activity can be useful in unwinding longer DNA fragments. The need of this activity can rise from the more constrained or even positively supercoiled DNA structure, which is believed to occur in the thermophilic organisms (Grogan, 1998).

On the other hand, really tight connection between the putative helicases and DNA repair proteins is observed in cases of the physical proximity of the methyltransferases and glycosylases (Fig. 23). Therefore, it is more likely, that the protein has some activity on the respective mRNA molecules. The helicase can possibly be involved in translation, for the efficient regulation of the translation termination of the "upstream" methyltransferase gene and the initiation of the "downstream" glycosylase synthesis (de la Cruz *et.al*., 1999). The importance of the well-balanced co-translation system is obvious from the tight functional coupling between DNA methylation and damage repair events. This could suggest a need of the helicase activity in the pass-by of the frameshift during protein translation. Alternatively, RNA editing events may possibly require the structures unwinding activity (Brennicke *et.al*., 1999). Finally, it was reported that DEAD-box proteins are needed to stabilise mRNA (Iost and Dreyfus, 1994).

4.4 Characterization of MutY.*Tth* from *Thermus thermophilus*

Early searches of the uracil-removing activity in the genome of *Thermus thermophilus* failed to reveal *orf* with significant sequence similarity to known glycosylases. At that time the only *orf*, with strong similarity to the helix-hairpin-helix class of glycosylases was identified. The respective *orf* shares the highest similarity to MutY of *E. coli* (about 40% identical and 56% similar residues) (Fig. 1.). According this, it was named MutY.*Tth*. However, more than 30% of identity and about 50% of similarity is preserved to the Mig subtype of the enzymes. Thus, the respective *orf* was cloned with the purpose of testing its activity for repair of U/G, T/G mismatches.

4.4.1 Cloning of the MutY.*Tth* of *Thermus thermophilus*

First, the sequencing data of the respective contig were edited. The region of interest had redundancy of 3-4, what quarantined the reliability of the data. The nucleotide sequence was then translated into the protein sequence in all six frames by the program DNA Strider. The a.a. sequence of interest was found, the start (*GTG*) and stop (*TAG*) codons were also identified. Based on that, the pair of primers was designed with the sites for *Nco*I and *Hind*III restriction endonucleases.

The product of the desired length was isolated by the PCR from the genomic DNA of *Thermus thermophilus*. Inclusion of the 5% DMSO was necessary for the obtaining of the product, most likely due to facilitated melting of secondary DNA structures. The reaction was performed with the *Taq* polymerase. After normal cycling profile, the reaction was incubated at 72^{o}C for 20 min to ensure extension of the 3' of the PCR product by A base (Dieffenbach and Dveksler, 1995). Then, the PCR product was cloned into the pCR^{R}2.1-TOPO vector via preformed T overhang. The resulting construct is shown in the Fig. 24.

The screen of the positive clones was done by the PCR (**3.2.14.1**). The protocol is modified in that way that the screening was done not from the single colony, but from the liquid culture, thus can be called “culture PCR”. The reason to modify the original protocol was due to the problems in processing the genes from *T. thermophilus.* This microorganism contains more than 70% of GC basepairs in the genome. In general, PCR reactions on such templates are inefficient because the DNA is prone to form stable secondary structures and DNA polymerase processes them with clearly reduced efficiency (Dieffenbach and Dveksler, 1995). In addition, “colony PCR” frequently inhibits the amplification due to the different cell components present in the reaction. Thus, a combination of these two factors resulted

often in false negatives. Analogous PCR on the liquid culture works more efficiently. Therefore, in course of the cloning of the genes from *T. thermophilus*, the modified protocol was used.

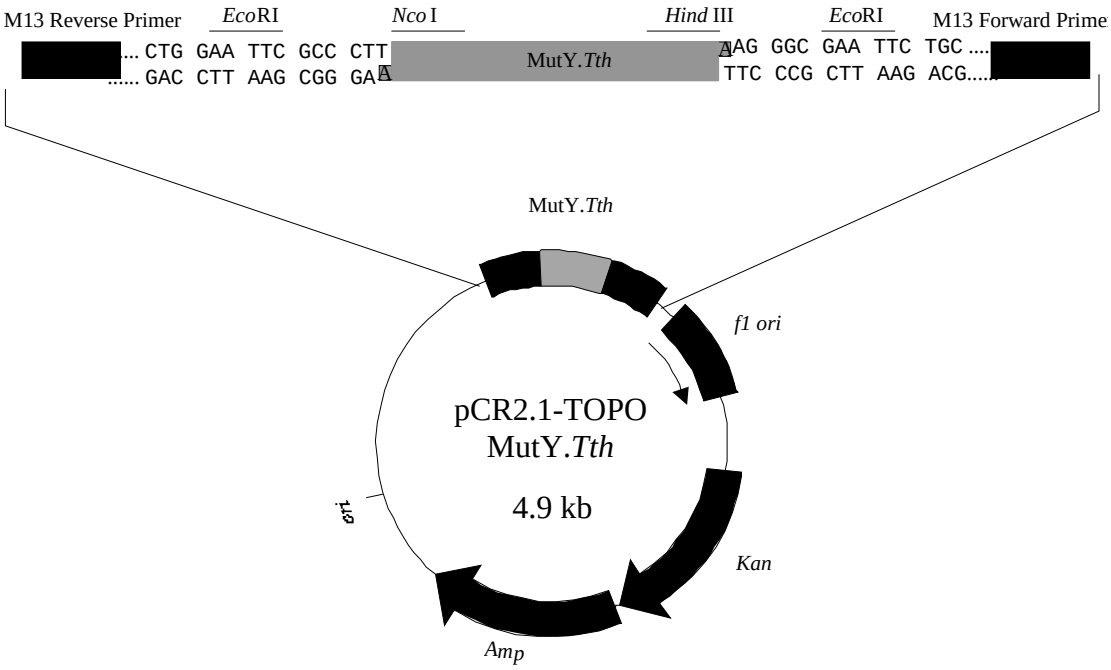


Fig. 24. Schematic representation of the cloning vector pCR 2.1 – TOPO/MutY.*Tth*. *ori* – ColE1 replication start of *Escherichia coli*. *f1 ori* – replication start of bacteriophage f1. *Amp* - β-lactamase gene. *Kan* – gene of the aminoglycoside phosphotransferase.

After the positive clones were confirmed by the DNA sequencing, the insert was re-cloned into the expression vector pET-21d. The resulting construct is analogous to that shown in the Fig. 5.

4.4.2 Expression and purification of MutY.*Tth* glycosylase

The plasmid pET-21d/MutY.*Tth* was transformed into the BL21(DE3)pLysS cells. The expression test was performed in the analogous way as described for the Mig glycosylases (**4.1.5.1**). In 2-3 hours after the induction with the 1mM IPTG, the band representing 38 kDa of the target protein was obtained in the soluble fraction

(data not shown). Then, further purification of the protein was done in the analogous steps to the Mig.*Mth*II (Fig. 25).

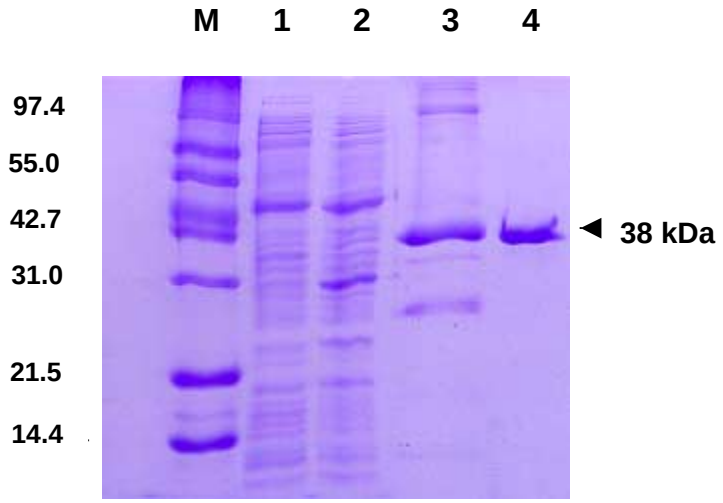


Fig. 25. Purification of the MutY.*Tth* protein. M - Mid – Range protein molecular weight marker (Promega). 1 – cells before induction of protein synthesis, 2 – cleared cell lysate after induction, 3 - fraction eluted from Ni^{+2} charged column with 200 mM imidazole, 4 – fraction eluted from HE POROS with 0.7 – 0.8 M NaCl.

As it is seen from the Fig. 26, the final protein preparation contains no visible bands of contaminating proteins. The MutY.*Tth* was diluted with 50% glycerol and stored at -20^{o}C.

Unfortunately, the yield of the protein is only 0.5-0.7 mg from 1l of culture. This is approximately 6-7 times less than amount of Mig.*Mth*II, despite the same vector and host used.

The wavelength scan resulted in the analogous picture as it was obtained for the Mig glycosylase (Fig. 10), thus, indicating of the presence of the [4Fe-4S] cluster.

4.4.3 Optimisation steps for the activity test of MutY.*Tth*

MutY.*Tth* has a high similarity to the MutY enzymes, for the initial activity tests A/G mismatch was chosen. Under the standard assay conditions (3.4.1), only marginal cleavage activity was obtained (data not shown). Therefore, some further experiments were devoted to optimise the assay conditions for this enzyme.

4.4.3.1 Incubation time

Since MutY glycosylases known to be slow (David and Williams, 1998), initial experiments were done to find optimal incubation time of the reaction (Fig. 26).

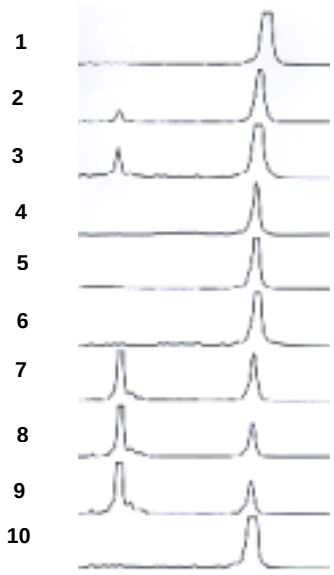


Fig. 26. Influence of the incubation time on the activity of MutY.*Tth*. **1, 2, 3** - reaction performed with the enzyme, **4, 5, 6** - without the enzyme for 20, 60 and 90 min, respectively. **7, 8, 9** – incubation of 30 U of MutY from *E. coli* (Trevigen) for 20, 60 and 90 min at 37 ^{0}C. **10** – negative control incubated at 37. A/G mismatch was used as a substrate.

It is obvious that prolongation of the incubation time noticeable increased the cleavage of the A/G mismatch, but the activity is still much lower than that obtained for the MutY.*Eco*. Thus, MutY.*Tth* acts as extremely slow enzyme, at least *in vitro* conditions.

4.4.3.2 Influence of different buffers and metal ions

The next step was to vary composition of the reaction. The set of buffers was tested:

1. 20 mM Tris-HCl, pH 7.6, 80 mM NaCl
2. 50 mM Tris-HCl, pH 9.0, 20 mM $(NH_4)_2SO_4$,

3. 20 mM Tris-HCl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2mM $MgSO_4$
4. 10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$
5. 33 mM Tris-acetate , pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate

Again, no significant change of enzymatic activity was noticed despite the differences in pH, concentration of the monovalent and divalent salts and buffering components. The buffer (1) is supplied together with the commercially available MutY.*Eco*, but no positive effect was noticed on the MutY.*Tth*. Also, parallel experiments with the MutY.*Eco* showed that the latter enzyme is also insensitive to the reaction composition. For the further measurements, buffer (I) was used as a standard buffer. The incubation time was prolonged to 90 min, while the temperature and other parameters remained unchanged.

4.4.4 Test of different mismatches

A set of 25 mismatches (**4.1.14**) was tested with MutY.*Tth* under the conditions indicated above. Unfortunately, the only mismatch cleaved was A/G. No traces of T/G or U/G cleavage were detectable. Also no activity on the ss DNA was obtained. The product peak was appearing only after heat-alkali treatment. Thus, the enzyme double strand specific, A/G DNA glycosylase. Is the enzyme indeed mono-functional glycosylase or no AP/lyase activity was detected due to the very little efficiency under the experimental conditions, remains unknown.

4.4.5 A/G cleavage in the different sequence context

The effect of sequence context on the activity of the enzyme was tested using A/G mismatch in two different sequence contexts (Fig. 27). The experiment revealed that, in general, the enzyme is non-discriminative towards sequence context. A similar situation was obtained for the MutY.*Eco*. However, it is known that the efficiency of the latter depends on the sequence context (Williams and David, 1998). Unfortunately, no refined data has been published (David and Williams, 1998).

The analogous test was performed with T/G and U/G mismatches. No activity was detected in any of the sequence surrounding; the same can be said about AT-rich substrates (data not shown).

Despite the efforts to optimise the assay conditions for MutY.*Tth* glycosylase, the enzyme cleaved only A/G mismatch. Thus the enzyme carries not only significant similarity to the MutY on the protein sequence level, but also exposes similar

behaviour. The motives involved into the substrate recognition "SEVLLQQT" and "NQALM" are well conserved between both enzymes. Also both glycosylases contain characteristic C-terminal sequence, needed for the recognition of the opposite G (Noll *et.al.*, 1999). Unambiguously, the MutY.*Tth* is a member of the MutY-subfamily.

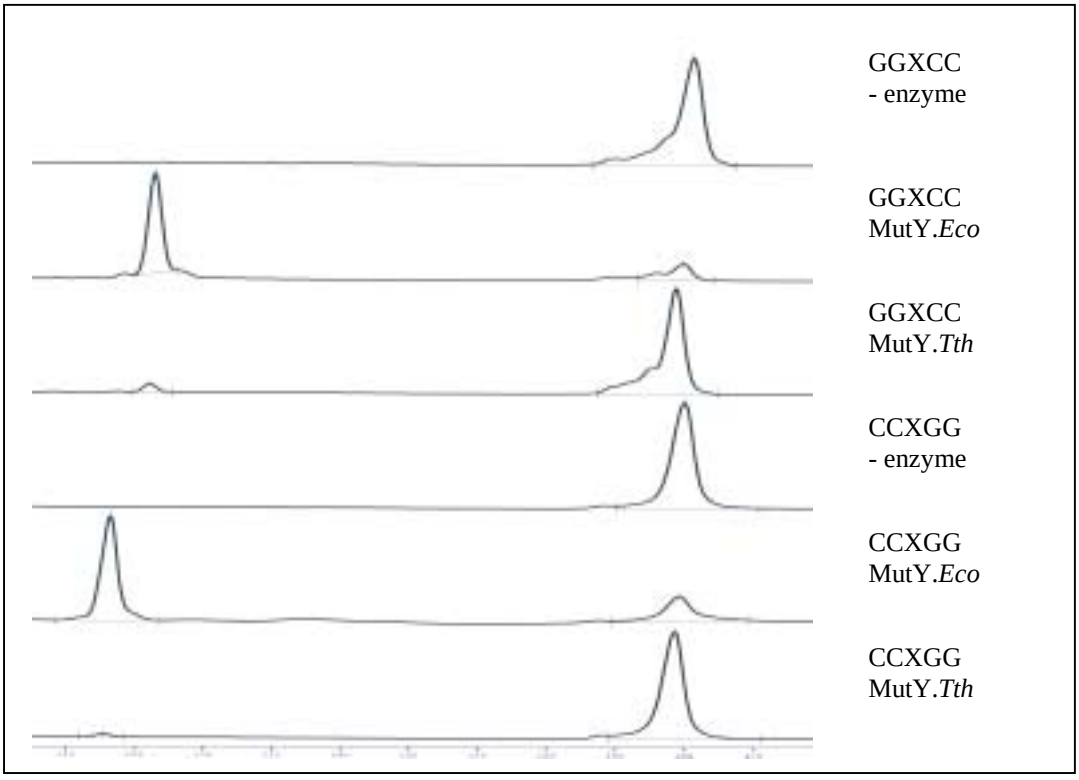


Fig. 27. A/G mismatch cleavage in the different sequence contexts by MutY.*Tth*. X-A/G mismatch.

MutY of *E.coli* cleaves A/oxoG mismatches with significantly higher efficiency than A/G or A/C mismatches (Williams and David, 1998). Thus, since MutY.*Tth* possesses only the activity towards A/G mismatch, it is possible that this is not its optimal substrate. Unfortunately, the base oppositions with oxoG were not tested.

Another possibility to explain little activity of MutY.*Tth* is, that the enzyme could require accessory proteins. This was described for the NthIII glycosylase (Klungland *et.al.*, 1999). The latter enzyme alone has only marginal activity towards oxidised pyrimidines, but becomes highly processive by the interaction with XPG – one of the components of NER machinery. Recent work on human MutY revealed

that the enzyme is stimulated by the interaction with AP-endonuclease (Yang *et.al*., 2001). Similar behaviour was noticed for the OGG1 glycosylase (Hill *et.al*., 2001) and hSMUG (Schärer and Jiricny 2001).

4.5 Characterization of the uracil glycosylases of *Thermus thermophilus* HB27

4.5.1 Searches of uracil glycosylases in the genome of *Thermus thermophilus*

During this project, the genome of highly thermostable bacteria *T. thermophilus* represented constantly growing and changing database of DNA sequencing data. Therefore, the homology searches of known uracil glycosylases were performed continuously. Initially, only one blast hit was identified. Unfortunately, as described in section 4.4., isolation the respective gene and characterization of the protein, named MutY.*Tth* showed the presence only of A/G, but not a uracil cleavage activity. With the growth of *T. thermophilus* database, and the description of novel uracil-glycosylase activities in 1999, two additional blast hits were identified. They revealed the presence of similar sequences to novel type of TMUDG glycosylase (Sandigursky and Franklin, 1999).

Work on these putative proteins started from editing of the raw sequencing data of the respective contigs. This was done in the Staden Package, GAP 4 subprogram. The base confidence reached was 99, with the error rate of 1/25000. The polished sequences were used to calculate the consensus sequence. This was then imported into NIP 4 subprogram and translated into the protein sequence in all six frames. The sequences of interest were found, stop and start codons identified. The *orf*s coding for the putative proteins were named TTUDGA and TTUDGB.

4.5.2 Analysis of sequences of TTUDGA and TTUDGB

The protein sequences of both were aligned to each other and to putative homologs from different organisms completely sequenced and annotated to date (Fig. 28). The multiple sequence alignment revealed that this class of glycosylase is widely distributed among various thermophilic and mesophilic organisms, both bacteria and archaea. Remarkably, no eucaryotic representatives were found. As seen from Fig. 29, the putative proteins exhibit very high degree of conservation with an average of 30 – 50% identity and 50 – 60% similarity. Majority of the microorganisms contain only one copy of the respective glycosylase. However, *T. thermophilus*, *M. loti* (Kaneko *et.al*., 2000) and *C. crescentus* (Nierman *et.al*., 2001) have two genes (indicated as A and B).

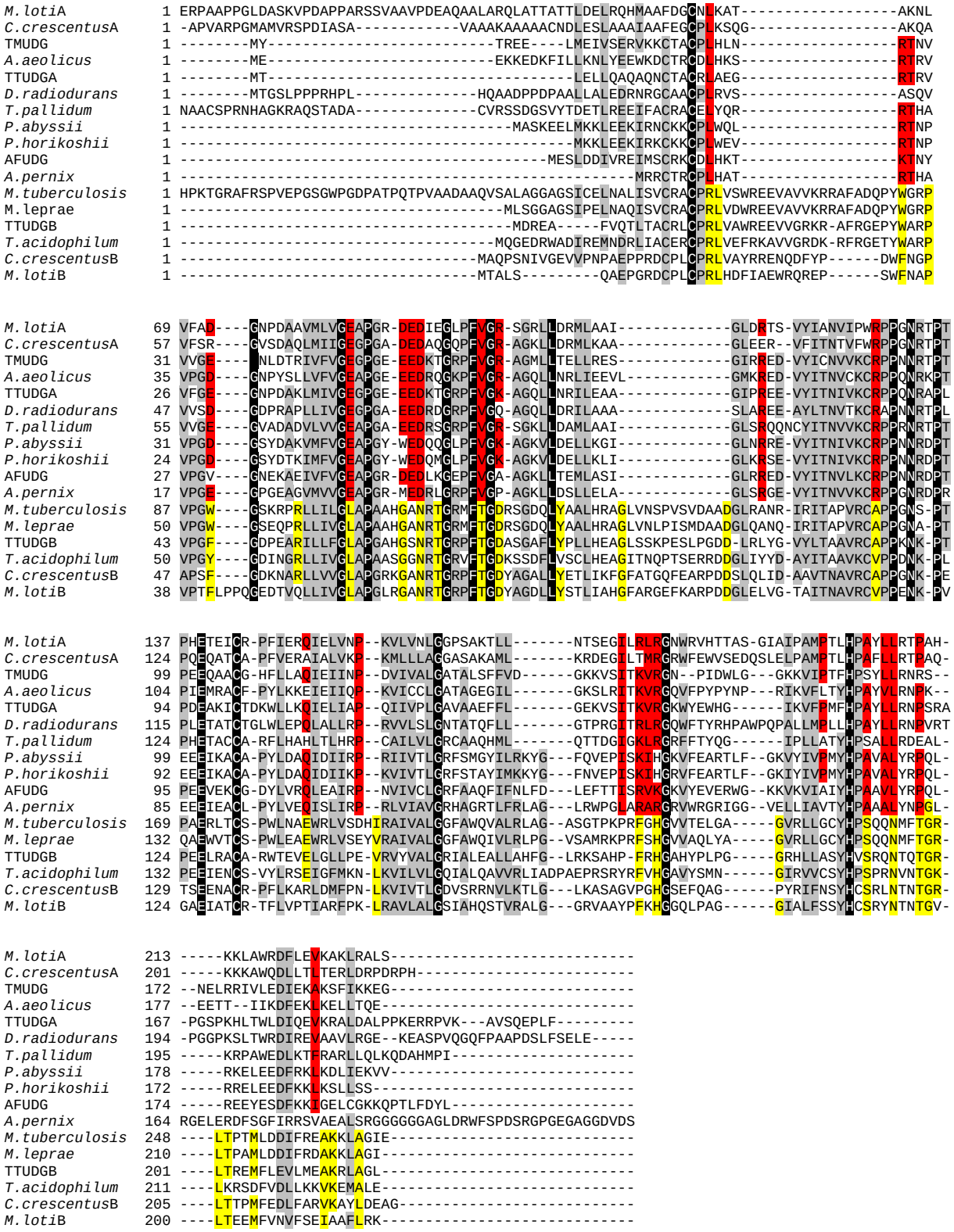


Fig. 28. Multiple sequence alignment of DNA uracil glycosylases of TMUDG family. Black color indicates conserved residues, gray - similar substitutions in all proteins. Red color indicates residues, which are characteristic only for A subtype, yellow - for the B subtype of these glycosylases. Full names of microorganisms are listed in the Fig.29.

Interestingly, in all three microorganisms genes for type A and B proteins are more similar to the sequences in other organisms than to each other (Fig. 29). They are, most likely then, obtained by different ways than gene duplication. TTUDGA glycosylase has the highest similarity (50% identity, 58% similarity) to a putative protein from *D. radiodurans* due to the close phylogenetic relationships (White *et.al*., 1999). High conservation (45% identity, 53% similarity) is also found to A type proteins from *M. loti* and *C. crescentus*: organisms with double copies of the respective genes. Thermostable representatives, like TMUDG (Sandigursky and Franklin, 1999) and AFUDG (Sandigursky and Franklin, 2000) also share significant degree of conservation (more that 40% identity, 50% similarity). As mentioned before, lower degree of similarity is observed to TTUDGB version (about 30% identity and 50% homology). The latter, on the other hand, has the closest sequences (50% identity, 58% similarity) in the genomes of *M. leprae* (Cole *et.al*., 2001) and *M. tuberculosis* (Cole *et.al*., 1998), also to the type B glycosylase residing in *M. loti* and *C. crescentus* (45% identity, 50% similarity).

All the proteins presented in the multiple sequence alignment have approximately similar length of about 200 - 230 residues; 16 of them are completely conserved in all sequences (shaded black in Fig. 28). About half (shaded grey in Fig. 28) are similar, thus, most likely, overall scaffold of the proteins is also similar. However, pronounced differences can also be identified. First, B subtype has two invariable conserved insertions. One of them is located close to the N-terminal part of the proteins and is 12-18 amino acids in length; the second is approximately in the middle of the sequences and carries 13 residues. The sequence within the insertions also bears significant similarity. Noticeably, both insertions have highly conserved boundaries. Second, there are residues specific only for the A or B subgroup (red or yellow colour in the Fig. 28, respectively). No intermediate proteins with mixed features were found.

Third difference appears after detailed comparison of the latter proteins with Ung and MUG glycosylases. These proteins have been crystallised and their structures solved (Savva *et.al*., 1995, Barrett *et.al*., 1998). Surprisingly, both glycosylases have a similar fold despite negligible conservation at the level of a.a. This made it possible to put them into one superfamily. In addition, Ung and MUG share several conserved motives, which are responsible for the interaction with uracil and cleavage of the glycosidic bond. Structure prediction was done for novel SMUG and TMUDG (A-type) glycosylases and the respective putative catalytic motives could be also identified (Pearl, 2000). Analysis of protein sequences of B type glycosylases also revealed the presence of two characteristic motives (Fig. 30).

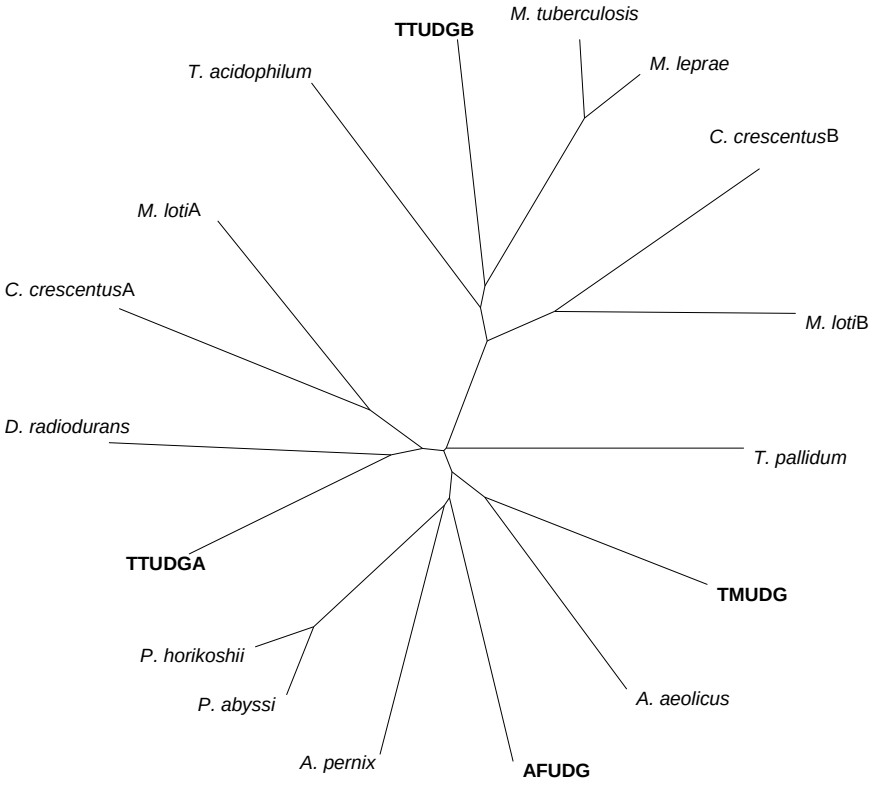


Fig 29. Phylogenetic tree of the glycosylases of TMUDG family. *A. aeolicus – Aquifex aeolicus*, AFUDG – *Achaeglobus fulgidus, A. pernix – Aeropyrum pernix, C. crescentus – Caulobacter crescentus, D. radiodurans - Deinococcus radiodurans, P. abysii – Pyrococcus abyssi, P. horikoshii – Pyrococcus horikoshii, M. leprae – Mycobacterium leprae, M. loti – Mesorhizobium loti, M. tuberculosis – Mycobacterium tuberculosis, T. acidophilum – Thermoplasma acidophilum*, TMUDG – *Thermotoga maritima, T. pallidum- Treponema pallidum*, TTUDG – *Thermus thermophilus*. Marked names represent experimentally tested activities.

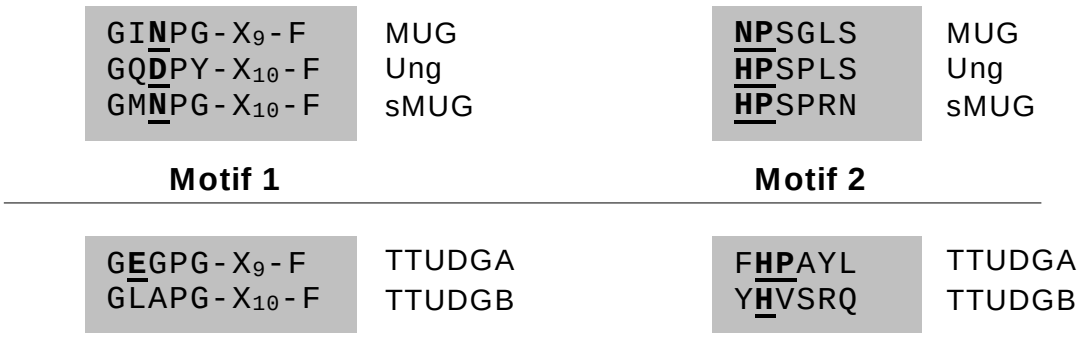


Fig. 30. Schematic representation of the conserved motives in the sequences of various uracil glycosylases. Sequences of MUG and Ung are from *E. coli* (accession numbers P43342 and P12295, respectively), sMUG – from *X. laevis* (AAD17300).

Shortly, motif I is known to make hydrogen bond interactions with flipped out uracil and also activates water molecule for the attack of glycosidic bond (review see Pearl, 2000; Schärer and Jiricny, 2001). Highly conserved phenylalanine located some nine or ten amino acids away stacks up the uracil base in the pocket. Catalysis, most probably, is performed by asparagine or aspartate residues in the third position of the motif (Fig. 30, Fig. 31). In case of TTUDGA this position is occupied by glycine (*Gly* 42), which is not able to activate water. However, this function, possibly, could be performed by the neighbouring glutamate (*Glu* 41). Surprisingly, B type of the enzyme does not contain any catalytic residue in the respective positions (*Leu* 58 and *Ala* 59). However, water activation possible could be performed by one of the further located aspartate residues (*Asp* 75 or *Asp* 101), which nevertheless could be “looped” in close proximity to glycosidic bond. Interestingly, the latter positions are conserved only among B enzymes (Fig. 28).

Motif 2 contains invariably conserved His (Fig. 30). It is widely accepted (for review see Barret *et.al*., 1999), that the residue makes hydrogen bond with O2 carbonyl group of uracil and, consequently, facilitates cleavage of glycosidic bond (Fig. 31). However, Mol *et.al*., 1995 proposed an alternative mechanism, where His (in case of TTUDGB, *His* 190) could play the major role in catalysis. Then, TTUDGB would be proficient in catalytic residue. Alternatively, TTUDGB might implicate completely novel mechanism of action.

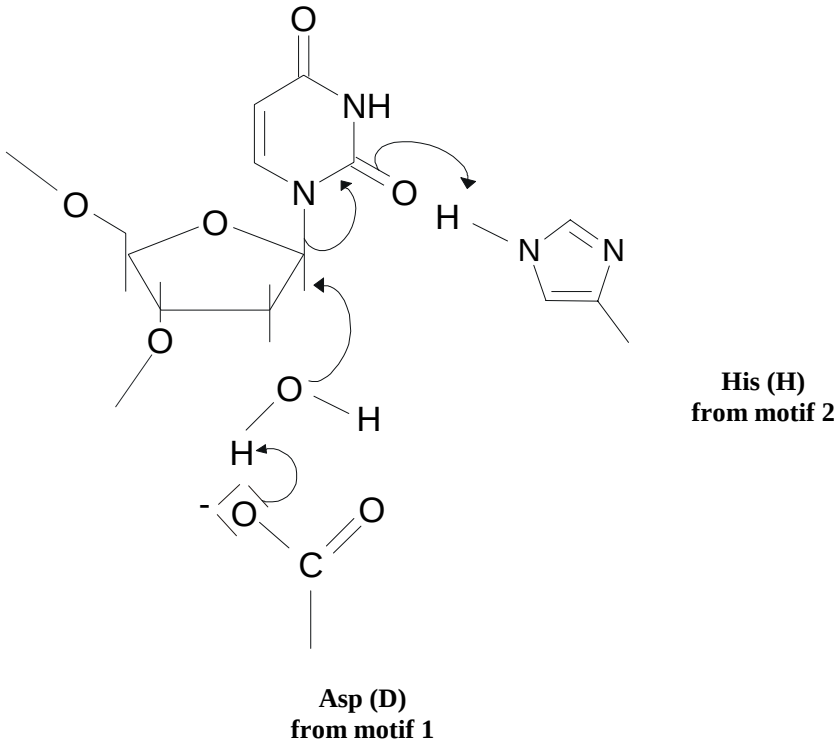


Fig. 31. Schematic representation of action of residues involved in cleavage of glycosidic bond by uracil glycosylases. Sequences of motives 1 and 2 are shown in Fig. 30.

4.5.3 Cloning the genes of TTUDGA and TTUDGB

The *orfs* coding for the TTUDGA and TTUDGB were isolated by PCR from the genomic DNA of *Thermus thermophilus*. The annealing temperature of 60 - 65 ^{o}C was used with DMSO obligatory present in the reaction mix. The reaction for TTUDGA was carried out with the primers containing *Nco*I and *Xho*I restriction sites. The primers specific for the TTUDGB contain *Nco*I and *Eag*I sites, respectively. Due to the extremely high GC content at the 3' end of TTUDGB gene and difficulties in obtaining PCR product, additional primers were designed. One set is analogous to the above mentioned; just do not contain sites for *Nco*I and *Eag*I restriction endonucleases (see Materials). Besides that, additional primer downstream of 3' end of TTUDGB (containing less number of GC nucleotides) was used (TTUDGBfar). PCR products were obtained with all primer pairs. However, only the targeted products were chosen for further experiments. Initially, PCR products were cloned into pCR Blunt II -TOPO vector. The correctness of inserts was confirmed by DNA sequencing. Then, the genes of TTUDGA and TTUDGB were re-cloned into the

expression vector pET-21d. The resulting constructs are of identical organization outlined in the Fig. 5. Due to the incorporation of the appropriate restriction sites, some changes were introduced into the protein sequences. In case of TTUDGA, the second threonine (*Thr*) is exchanged into the alanine (*Ala*), the last phenylalanine (*Phe*) – into the glutamine (*Glu*). For the TTUDGB, the N-terminal part was left without changes, while the additional *Ser Ala Ala Leu Glu* string was introduced at the C-terminal part after the last unchanged leucine (*Leu*) residue. Both constructs utilize "*ATG*" as a start codon and contain hexahistidine tails on the C-terminal ends.

4.5.4 Expression and purification of TTUDGA and TTUDGB glycosylases

The plasmids pET-21d/TTUDGA and pET-21d/TTUDGB were transformed into the BL21(DE3)pLysS cells. In general, the scheme of the proteins expression and purification was similar to that of Mig.*Mth*II glycosylase (4.1.5). Shortly, the cells were grown at 37 ^{o}C till OD_{600} reached 0.6-0.7. Then, the protein synthesis was started by addition of 1mM IPTG. After 3 hours of incubation cells were collected by centrifugation and lysed. The protein was purified by the two subsequent chromatographic steps: initially by an IMACS column, then followed by a HS 20 column on a Biocad Workstation (Fig. 32). All steps of protein purification were performed in the buffer 25 mM HEPES, pH 7,6.

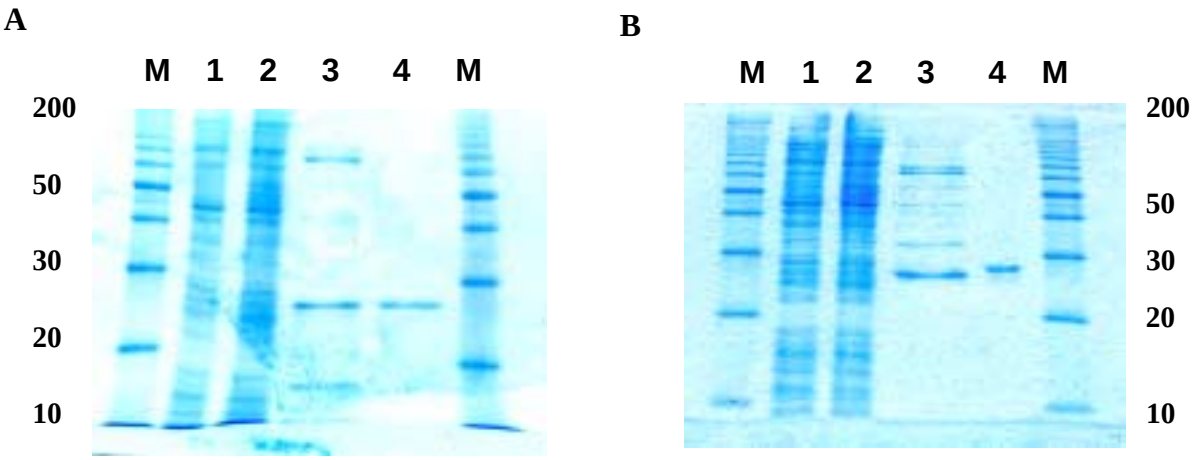


Fig. 32. Purification scheme of TTUDGA **(A)** and TTUDGB **(B)**. **M:** 10 kDa protein size marker (Gibco), **1**: crude extract of uninduced cells, **2**: crude extract of IPTG induced cells, **3**: protein, eluted with 300 mM imidazole from IMAC column, **4**: protein, eluted with 0.5 – 0.6 M NaCl from HS 20 column.

As seen from the Fig. 32, the purification scheme leads to the apparently homogenous proteins: TTUDGA has the molecular weight of 23.7 kDa, TTUDGB - 25.2 kDa, respectively. Following protein concentration was done by the Centriprep filter with the cut-off value 3000. For the long-time storage, both glycosylases was diluted with 50% glycerol and placed into -20^0C.

Thus, TTUDGA and TTUDGB can be expressed and purified in a similar way as the helix-hairpin-helix glycosylases. However, there are several pronounced differences in the protein behavior. First, the yield of the protein production is about 10 times lower than that obtained for the Mig enzymes from *M. thermoautotrophicum*. Only 0.2-0.3 mg of the TTUDGA and TTUDGB was produced from 1 liter of the culture, despite the similar expression conditions and the same vector-host combination used. Little expression level was observed also for the other protein from *T. thermophilus* – MutY.*Tth* (4.4.2). Therefore, the low expression levels, most likely, appear due to the codon usage differences between *E. coli* and *T. thermophilus* cells (Dr.R.Merkl, personal communication). The type of glycosylase, apparently, does not play an important role.

Other variation in the purification scheme originates from the intrinsic properties of proteins from *T. thermophilus*. Usually, DNA mimicking polymer heparin can be used for the purification of DNA binding proteins (Farooqui, 1980). This was successfully applied for the purification of helix-hairpin-helix class of glycosylases (4.1.5.2). Surprisingly, TTUDGA binding to heparin column was inefficient: part of the protein was detected in the flow-through. In addition, not all contaminants were removed from the fractions eluted with NaCl gradient (data not shown). Then, different columns were tested and the strong cation exchanger HS (sulphopropyl) was found to be suitable. The latter column attracts positively charged molecules. As the purification buffer has pH value of 7.6, TTUDGA and, most likely, TTUDGB behave as cations (pI = 9,3 for TTUDGA, pI = 10.5 for TTUDGB, respectively; calculated by GCG Sequence Analysis Software Package). This, apparently, ensures the efficient separation.

4.5.5 Initial characterisation of TTUDGA and TTUDGB as uracil-glycosylases

TTUDGA and TTUDGB putative glycosylases were expressed in the ung and mug proficient host, therefore, the proteins were preheated at 70 ^{o}C for 15 min before every cleavage reaction. Knowing that Ung enzyme in *E. coli* cell is fully inhibited already at 60 ^{o}C (Lindahl, 1977), interference with the activity of the recombinant

proteins of *T. thermophilus* is very unlikely. The activity assays were performed with U/G mismatch embedded in the sequence context "GGXCC", as well as with the single stranded substrate carrying the analogous sequence. U/A mispair was surrounded by "CCXGG". The reaction conditions used were as described in the **Methods** section, with the reaction buffer of pH value 9.0 (room temp.). The assay results are presented in Fig. 33.

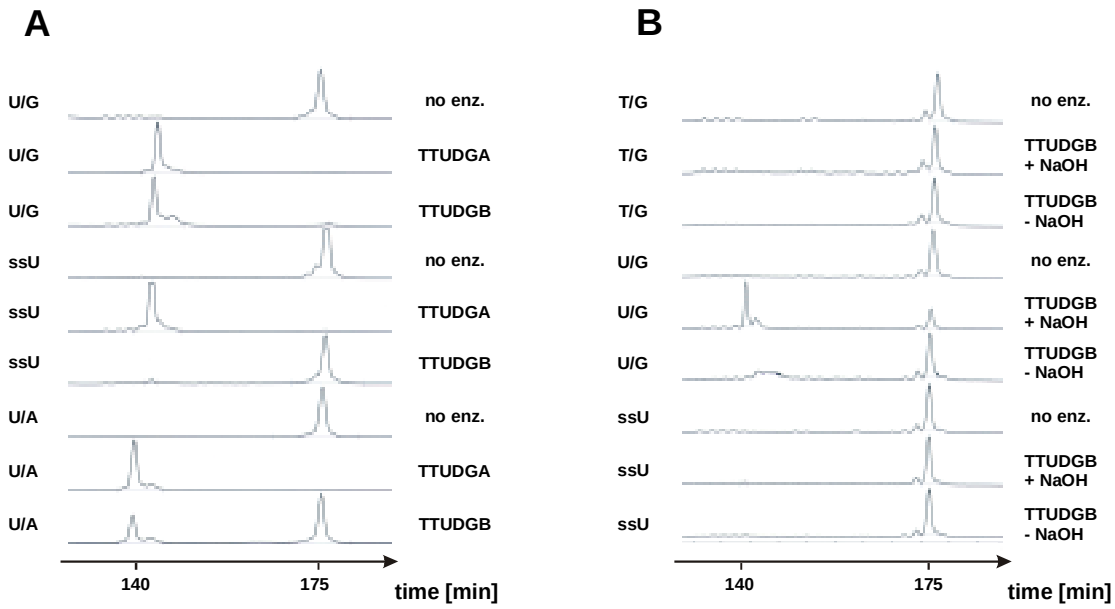


Fig. 33. Activity test for presence of uracil glycosylase activity in TTUDGA and TTUDGB. Synthetic DNA duplexes used in the assays are the same as showed in Fig. 11. A – two-step assay of glycosylase activity by TTUDGA and TTUDGB, B – presence of AP-lyase activity and test of activity towards T/G mispairs and by TTUDGB.

As seen from the Fig. 33 A, both enzymes are uracil glycosylases. They are able to remove U/G mismatch with the comparable efficiency. However, ability to process uracil in the ssDNA differs dramatically. TTUDGA acts in seemingly efficient way, while TTUDGB activity is marginal. Pronounced differences are also in cleavage of U/A: despite that TTUDGB enzyme exhibits some activity, this is, however, much lower than that of TTUDGA. No T/G cleavage was detected by TTUDGB (Fig. 33 B). The same is true for TTUDGA (data not shown). Experiments without addition of NaOH and heat – treatment, revealed that both enzymes have no associated AP –

lyase activity nor on ds, neither on ss DNA (Fig. 32 B for TTUDGB, for TTUDGA data not shown).

Thus, the test showed that the TTUDGA, as expected, is a typical monofunctional uracil glycosylase, strictly specific for uracil in ds and ss DNA. These data are in good agreement with the A type enzymes purified from *Thermotoga maritima* TMUDG (Sandigursky and Franklin, 1999) and *Archeoglobus fulgidus* AFUDG (Sandigursky and Franklin, 2000). Accordingly, TTUDGA could serve as uracil repair activity of broad action, thus, representing the functional homolog of Ung glycosylase.

TTUDGB enzyme, however, requires more complicated interpretation. In general, the presence of uracil - removal activity was surprising due to the fact that the enzyme might lack putative catalytic important amino acids (Fig. 30). In difference from TTUDGA, TTUDGB requires double stranded DNA. It is not completely ds DNA specific enzyme, as Mig.*Mth* glycosylases, but uracil is removed from ss template highly inefficiently (Fig. 33). Thus, TTUDGB represents the enzyme, whose substrate recognition and binding properties are different from already characterized uracil glycosylases. Consequently, the enzyme hardly duplicates the work of TTUDGA, thus its biological role could be very different.

4.5.6 Substrate requirements of TTUDGA and TTUDGB

To obtain the full spectrum of the substrate requirements of TTUDGA and TTUDGB, the set of different mismatches (4.1.14) was tested. The reaction conditions were as following: 40 fmol of fluorescently labeled substrate was mixed with the 10 - 12 pmol of the enzyme in the 50 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$ buffer, pH 9.0. The reaction was incubated for 30 min at 50 ^{o}C and subjected to the heat/alkali treatment. The following A.L.F. run showed that both enzymes remove uracil not only from U/G and U/A, but also from any other opposition in dsDNA, like U/C, U/T and U/U (data not shown). Other mismatches, like T/G, T/C, A/C, A/G, G/G, C/C were not processed. In addition, reaction buffer with pH 7.5 (room temp.) was tested, as well as influence of Mg^{+2} ions: no significant changes in activity of both enzymes were identified.

4.5.7 Activity on single stranded DNA substrates by TTUDGA and TTUDGB

Since cleavage of ssDNA represents the most striking difference between TTUDGA and TTUDGB, further, more detailed tests were performed. For this, reactions were composed with a different enzyme to substrate ratio, as well as concentrations. Also, influence of flanking DNA sequences was tested. Three ss DNA substrates of varying sequence patterns (TTUAA, GGUCC and TGUCT) were chosen. Each of the substrate was subjected to the assays with following composition of reaction: (1) 12 pmol enzyme and 40 fmol substrate, (2) 4 pmol enzyme and 400 fmol substrate, (3) 4 pmol enzyme and 4 pmol substrate, (4) 40 pmol enzyme and 4 pmol substrate, (5) 40 pmol enzyme and 400 fmol substrate. The reactions were performed at 50 ^{o}C, 30 min with TTUDGA and TTUDGB, separately. Results are summarized in Fig. 34.

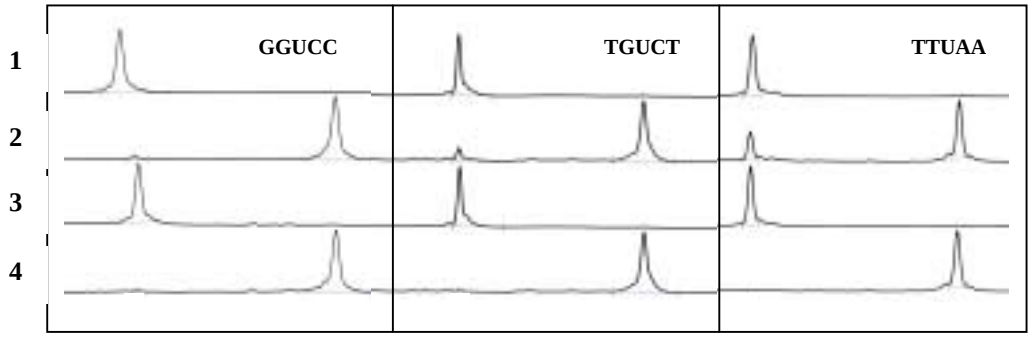


Fig. 34. Activity of TTUDGA and TTUDGB on single stranded DNA substrates. 1. 12 pmol of TTUDGA and 40 fmol of the respective substrate; 2. 12 pmol of TTUDGB and 40 fmol of substrate; 3. 4 pmol TTUDGA and 4 pmol of substrate; 4. 4 pmol TTUDGB and 4 pmol of substrate. Substrate with the sequence pattern GGUCC (MU2-9) consists of 39 bases, TGUCT (FYT*T) - 43, TTUAA (FYT*A) - 31. Full sequences of substrates are presented in Materials.

The latter experiment revealed further differences between tested proteins. TTUDGA glycosylase processed uracil in ss substrates very efficiently, irrespective of the sequence context and composition of the reaction. (1 and 3 lanes in Fig. 34). TTUDGB, being very inefficient on ssDNA in general, also showed quite strong requirements to particular sequence context. GC rich substrate was processed only marginally (the same as documented in Fig. 33), introduction of several T bases

increased the activity up to 10%, in AT rich surrounding about 30% of substrate was cleaved (lane 2 in Fig. 34). Unfortunately, in difference from TTUDGA enzyme, this activity could be detected only with huge excess of enzyme (no less than 100 times) in the reaction (1 and 5 compositions). If the components are in the equimolar parts or enzyme : substrate ratio at about 10 : 1 (2, 3 and 4 compositions) no activity is detectable (lane 4 in Fig. 34). Absolute concentrations of the reactants, apparently, had no influence.

Interpretation of the latter experiment is rather complicated at this stage of work. Marginal activity of TTUDGB enzyme on single stranded substrates could originate from not proper binding to ss DNA, thus the access to the glycosidic bond could be strongly hindered. Requirements for specific sequence context could reflect the conformational influence on proper orientation of uracil, hence favoring the suggestion. However, small amount of product on ss DNA becomes gradually still less upon increasing the assay temperature from 50 ^{o}C to 90 ^{o}C (data not shown). Therefore, the observed activity in some sequence context could originate from transient self-annealing of the substrate at lower temperatures. Certainly, clarification of this is highly welcome.

4.5.8 Optimization of the multiple substrate kinetics assay for TTUDGA and TTUDGB

As the next step in evaluation of enzymatic properties and mode of action of both proteins, the approach of the multiple substrate kinetics was applied (Schellenberger *et.al*., 1993). For TTUDGA, ds and ss DNA substrates, different only in their length (4.1.13) were mixed and processed under the standard conditions described in the Methods section. Analogous was done for TTUDGB, only with set of ds substrates.

The assay showed that standard assay conditions are not suitable for the work neither with TTUDGA nor TTUDGB. In case of A enzyme, both substrates ss and ds DNA were converted into the products within the first 10-20s, thus no quantitative measurements could be performed (Fig. 35, A). Therefore, next steps were directed to find optimal enzyme : substrate ratio, which would be suitable for the kinetic assay within the reasonable time period. For this, concentration of the TTUDGA was gradually reduced and, additionally, the concentrations of the substrates were

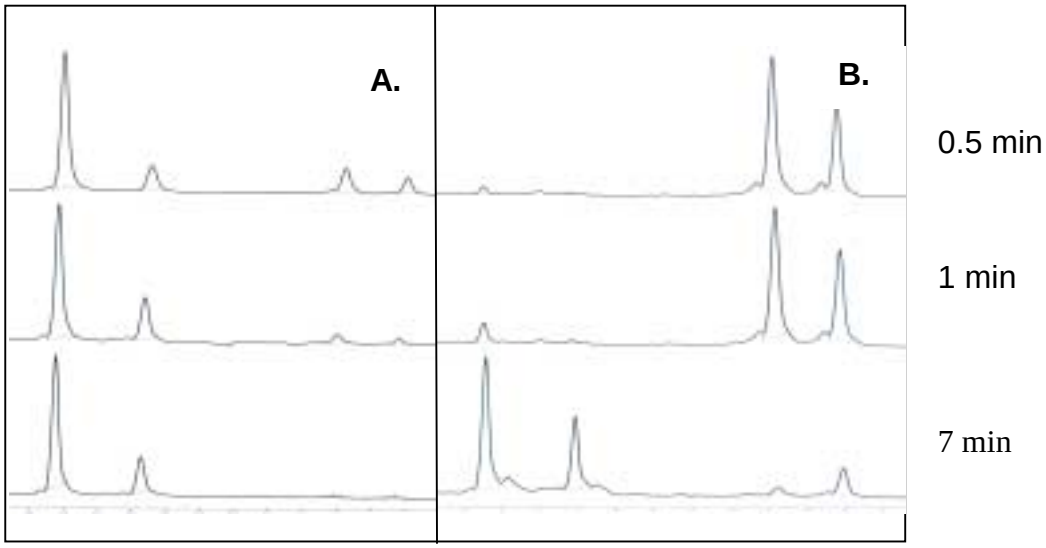


Fig. 35. Influence of the enzyme and substrate concentrations on the efficiency of the cleavage reaction by TTUDGA. **A:** initial conditions: 10 pmol of TTUDGA and 40 fmol each dsDNA substrate, **B**: end conditions: 1 pmol of TTUDGA and 20 pmol of each dsDNA substrate, consisting of 40 fmol labelled and 19.96 pmol "dark" DNA. U/G mismatch or uracil in ss DNA template were embedded in CCUGG sequence pattern. Shorter substrate consists of 35 bases (MU2-9), longer - of 39 (MU-3). Full sequences are presented in **Materials**.

increased. The limitations to the amount of fluorescently labelled substrate is dictated by the detection system of A.L.F. sequencer, thus the increase of the substrate concentration was achieved by the addition of analogous, but unlabelled, "dark" substrate. After several optimisation rounds, the assay conditions were found, which were suitable for both ss and ds DNA (Fig. 35, B).

The latter experiment demonstrated very high efficiency of TTUDGA enzyme, while overall 40 times excess of substrate was needed to slow down the reaction. Unfortunately, these conditions were not suitable for TTUDGB enzyme - no activity was detectable (data not shown). Then, opposite steps were taken, e.g. concentrations of substrates were gradually decreased. Concentration of protein, however, was kept the same as for TTUDGA (1 pmol). Cleavage rates similar to that obtained by TTUDGA were achieved with only 50 fmol of overall DNA concentration (data not shown), so addition of "dark" substrate was not necessary.

Thus, 1 pmol of TTUDGA completely cleaves 40 pmol DNA (1 : 40 ratio) in about 10 min time period. Within the same time, 1 pmol of TTUDGB enzyme

processes only 50 fmol DNA (20 : 1 ratio). So, resulting difference between activities of the enzymes is about 800 fold. Important, that comparison can be applied to the activity only on ds DNA, while ss DNA is processed in not measurable rates by TTUDGB (Fig. 33).

4.5.9 Mode of action of TTUDGA on ds and ss DNA

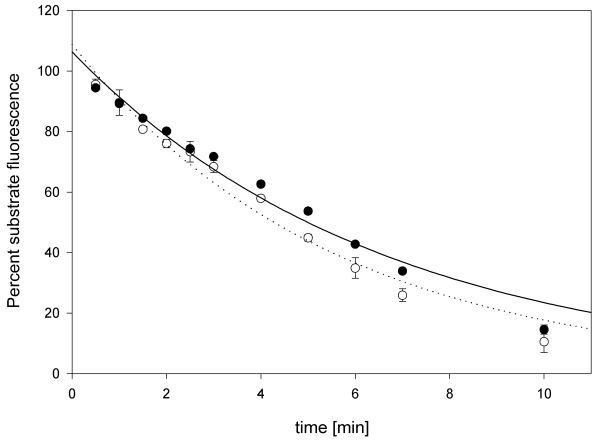
The scheme of above described experiment was used for further insight into enzymatic properties of the enzyme. Sensitivity of TTUDGA to the length of DNA was tested by integration and comparison of consumption rates of each substrate separately. As described earlier, the oligonucleotides used differ in 4 bases. Such studies were performed for ds and ss DNA, under optimised conditions (4.5.8) and results are presented in Fig. 36. As seen, no difference in activity of TTUDGA was brought in by the length of ds and ss DNA under conditions used. Thus, the kinetics assay with the mixture of substrates is applicable to the enzyme. As expected, ds and ss oligonucleotides are clearly processed with comparable rates, the latter being slightly more efficient.

In addition, the experiment served as quantitative illustration of high activity of TTUDGA. Thus, an efficient cleavage even huge excess of the DNA, substrate, most likely, indicates high turnover number of TTUDGA glycosylase.

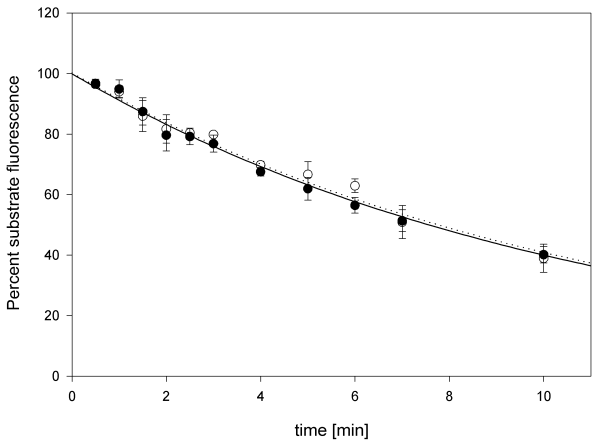
4.5.10 Efficiency of uracil removal from different mispairs by TTUDGA

As mentioned before, TTUDGA possess the ability to remove uracil from various base oppositions in ds DNA molecules (**4.5.6**). However, these qualitative measurements were done with huge excess of enzyme, what, obviously, hindered the specificity to a particular substrate. Thus, the influence of the opposite base on the efficiency of uracil removal was tested by multiple substrate kinetic assays under conditions described (**4.5.8**). The mixes of oligos carrying U/A, U/G and U/C base/base oppositions were measured, with U/G mismatch as a reference substrate. The mismatches were embedded in the CCUGG sequence context. All assays were performed in triplicate and values averaged.

The experiment revealed significant rates of the discrimination by TTUDGA dependent on the interaction partner of uracil (Fig. 37). The most easily uracil is removed from U/C base/base opposition, the most slowly - from U/A match with the discrimination rate at about 150 times. U/G mismatch is processed with the middle



Single strand DNA



Double strand DNA

Fig. 36. Influence of the substrate length on the activity of TTUDGA. Black circles indicate long, white – short DNA substrates.

rate. The observation is in good agreement with published data for the likely functional homolog of the TTUDGA - Ung glycosylase (Krokan *et.al.*, 1997). Despite that, no quantitative evaluation is available for the latter, the mode of substrate preference seems to be similar for both enzymes.

The observed discrimination against different oppositions of the uracil correlates well with the proposed mechanism of the action of Ung type uracil glycosylase (Parikh *et.al.*, 1998). It was determined that the enzyme acts via base flipping (Slupphaug, 1996). The latter facilitates cleavage of the glycosidic bond, by positioning the mismatch base in the extrahelical conformation. Efficiency of base flipping out of double stranded helix, is directly dependent on the strength of base/base interactions inside it. Accordingly, U/A is a normal match with the uracil stacked firmly inside the helix. Thus, the enzyme possibly requires the additional energy to breake the interacting bonds and, consequently, the cleavage of uracil is strongly hindered (Fig. 37). U/G mismatch is much less stable than U/A match, so increase in uracil removal rate is in the range of expectations. That is about 20 times different between U/A and U/G oppositions for TTUDGA. Finally, U/C mismatch is the least stable. In the agreement to that, the measured rate of the uracil cleavage is increased additional 7 times (Fig. 37).

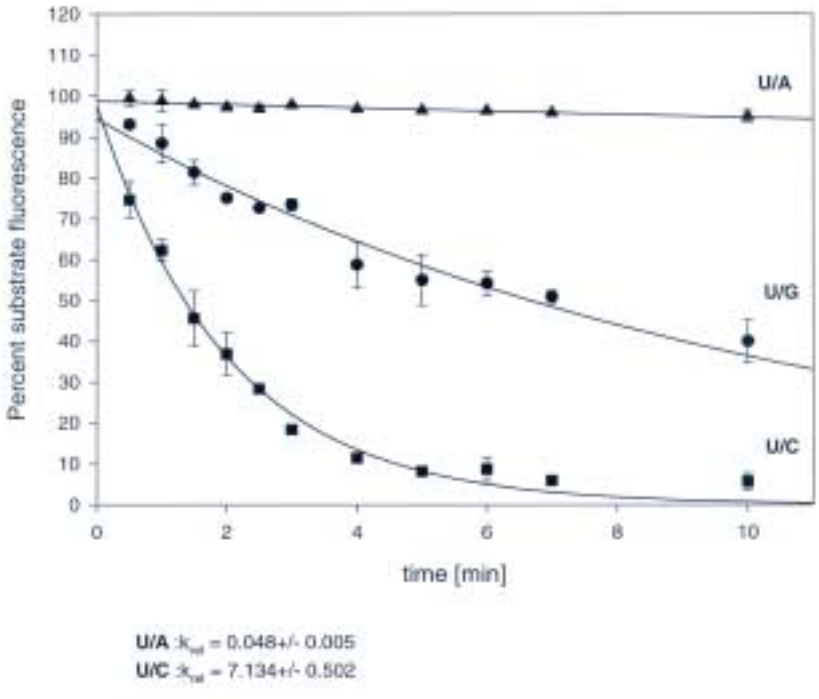


Fig. 37. Opposite base influence on uracil removal efficiency by TTUDGA.

Thus, the experiment revealed that TTUDGA glycosylase indeed follows similar substrate recognition and removal principals as described for the Ung enzymes (Slupphaug *et.al.*, 1996, Panayotou *et.al.*, 1998), despite negligible overall sequence

similarity. The result is not unexpected, while threading of the protein sequence of this class of the enzymes (Pearl, 2000) showed that the latter could have similar fold to MUG and Ung enzymes (Barrett *at.al*., 1998, Mol *et.al*., 1995). Consequently, similar catalytic action could be suggested. Also, similar to the Ung glycosylase, the rate of the uracil removal from the ss DNA is slightly higher than from duplex with U/G mismatch (Fig. 36). This was also explained by the fact that the cleavage of the unpaired uracil in the ss DNA requires less energy than the wobbling U/G mismatch (Panayotou *et.al*., 1998). However, the most efficient cleavage of U/C base/base opposition by TTUDGA, remains without explanation, unless it would be related to the preferential binding of the enzyme to ds DNA. Unfortunately, no data about the activity on U/C mispair are available for Ung enzyme.

4.5.11 Influence of the sequence context on the activity of TTUDGA

It was observed that the activity of Ung enzymes depends on the sequence context around the mismatched uracil (Krokan *et.al*., 1997). Especially pronounced rates were measured in the double stranded substrates, with the amplitude of the discrimination at about 15 times (Eftedal *et.al*., 1993). This is true both for U/A matches and U/G mismatches. The most efficient removal of the uracil was measured in the AT-rich sequences; the retarded repair was in the GC-rich regions. Consequently, the observations were explained by the facilitated melting of duplex DNA at AT-rich sites, what leads to a better accessibility of the uracil by the Ung glycosylase (Panayotou *et.al*., 1998). The rates of discrimination among single stranded substrates were also measured (Delort *et.al*., 1985) and found to be only 3 times different between the worst and the best sequence contexts. Therefore, the specificity of the Ung enzyme to the particular sequence pattern was initially excluded. However, later works (Nilsen *et.al*., 1995) suggested that the observed rates of the discrimination among ds substrates are too big to be explained only by the local melting. Especially, that the preference of ss over ds DNA by the Ung is only about 2-3 times (Domena and Mosbaugh, 1985). Thus, the sequence-dependent variations of the Ung activity remained not-fully clarified. Some authors (Nilsen *et.al*., 1995) associated it with the hotspots for the cytosine deamination in *E.coli*. The latter effect is more pronounced in the AT-rich contexts, hence, overlaps with the favorable substrates of the Ung glycosylase.

The influence of the sequence contexts upon the activity of TTUDGA glycosylase was also tested with the aim to obtain more information about the

possible function of the enzyme *in vivo*. The multiple substrates kinetic assay was performed with U/G containing ds and U containing ss DNA oligoes. The sequence $CC^{U}GG$ was considered as a reference substrate in all measurements. Set of the substrates with the varying degree of AT- and GC-bases was tested (Fig. 38). Measurements with ss substrates were done in triplicate and the values obtained were averaged, while ds substrates were measured only once, thus the data serve as a guideline, but not as a final result. The relative rate constants are summarized in the Table. 14. and compared with the data for the Ung enzyme.

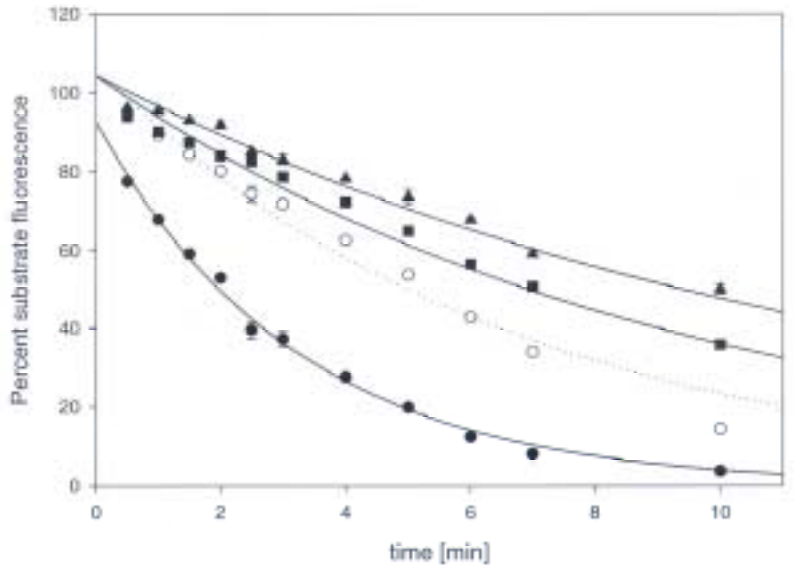


Fig. 38. Discrimination towards sequence context among ssDNA substrates by TTUDGA. White circles indicate the reference sequence CCUGG (MU2-9), black circles – sequence TTUAA (FYT*A), black squares – TGUCT (FYT*T), black triangles – GGUCC (FYGC). Assay with ds DNA substrates results in the similar picture. Full sequences of the substrates are presented in the **Materials**.

The experiment revealed the presence of easily detectable, but modest discrimination among various DNA sequences by TTUDGA glycosylase. In difference from Ung, only 5-6 times differences between the best and worst substrates were obtained. In addition, the performance and ranking on ss and ds substrates were the same (Table 14). Therefore, the observation hardly can be attributed to the local DNA melting, but rather represents the intrinsic properties of the enzyme.

Sequence	ss DNA by TTUDGA	ds DNA by TTUDGA	ds DNA by Ung[1]
TT^{U}AA	2.47+/-0.01	1.44	588%(100%)[2]
CC^{U}GG	1.00+/-0.00	1.00	100% (17%)
TG^{U}CT	0.57+/-0.02	0.61	153-100%(26-17%)
GG^{U}CC	0.41+/-0.03	0.25	<50%(8.5%)

[1]Eftedal *et.al*., 1993.

[2] Cleavage efficiency in the brackets showed according original description.

Table 14. Substrate specificity of the TTUDGA and Ung glycosylases. The specificity of the TTUDGA is expressed as relative rate constants, of the Ung – in percents of the uracil removal within the fixed time period.

Interesting, that the order of the substrate preference by TTUDGA is in a good agreement with the data of the bacterial Ung glycosylase (Eftedal *et.al*., 1993). For the latter, values only of the ds substrates are included, because the overall design of the ss substrates was very different (Delort *et.al*., 1985) and could not be compared with data for TTUDGA. In any case, however, the most favourable substrate is an AT rich sequence. The cleavage of the latter substrate is clearly preferential by Ung enzyme. The analogous DNA sequence is also cleaved efficiently by TTUDGA. The worst processed sequence by both enzymes contains stretches of GC. This result is rather surprising for TTUDGA, while more than 70% of the genome of *T.thermophilus* consists of GC bases and, consequently, GC rich sequences serve as the most abundant natural substrate.

4.5.12 Efficiency of uracil removal from different mispairs by TTUDGB

Qualitative activity test (4.5.5) revealed that TTUDGB glycosylase is able to remove uracil from various oppositions. More detailed mode of action was investigated by the means of multiple substrate kinetics under conditions optimised (4.5.8). As the initial step, measurements of the influence of substrates length were performed. Similar to TTUDGA, no influence was detected (data not shown). Thus, rates of cleavage of U/G, U/A and U/C were measured and data presented in Fig. 39.

The experiment revealed striking differences between TTUDGA and TTUDGB enzymes. The latter appeared not to be specific for uracil, but to U/G mismatch, while U/G is processed with significantly higher efficiency than U/A or U/C. This observation is in a good agreement with specificity of TTUDGB to ds DNA. Most likely, successful cleavage reaction requires interactions not only with uracil, but also with the orphan base. Similar performance was described for ds DNA specific MUG glycosylase (review Pearl, 2000). Thus, damage recognition process, possibly

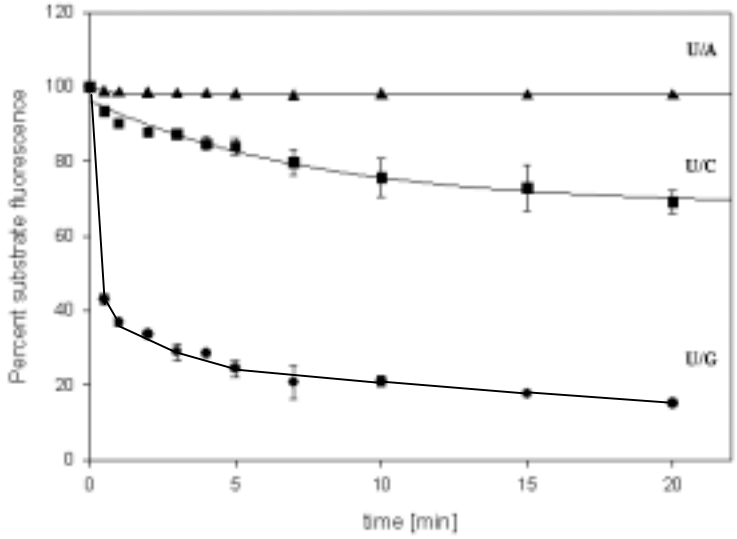


Fig. 39. Removal of uracil from different oppositions by TTUDGB. As U/G mismatch is processed by unusual way (more in the text), exponential curve fit was not performed. The points were joined by hand for visual reasons. Consequently, no relative rate constants are presented.

is primarily dictated by the specific interactions with G base rather influence of the strength of hydrogen bonds inside the DNA helix. However, TTUDGB might act in different way as MUG, because, in difference to the latter, cleavage of U/A and U/C was easily detectable. Thus, the interactions with opposite base are not completely exclusive. In addition, clear preference of U/C over U/A demonstrates possible influence of hydrogen bonding as in case of TTUDGA. Thus, TTUDGB apparently carries features of both uracil - and mismatch - specific glycosylases.

Another difference in performance of TTUDGB originates from the reaction composition and the shape of the activity curve for U/G cleavage (Fig. 39). In difference from TTUDGA, already after few minutes of incubation majority of substrate is converted. During further incubation, very slow reaction is still detectable. Interpretation of this observation is rather complex. Quick initial and slow following U/G cleavage, most likely, represents high catalytic activity and very strong binding of the enzyme to the product and/or the substrate. Thus, in difference from TTUDGA, TTUDGB can be characterised by low processivity. However, slow following reaction, indicates that the enzyme can still diffuse from the complex with DNA and, most, likely, initiate new catalytic event. The whole picture is further complicated by the fact, that TTUDGB must be present in huge excess in the reaction (4.5.8). Possibly, only small portion of the enzyme possess catalytic activity. Alternatively, additional processivity factors could be needed. Cleavage of U/C and U/A occurs more smoothly, what possibly could indicate weaker binding of TTUDGB to such DNA substrates. On the other hand, activity on U/C and U/A mispairs could be non-sufficient to detect step-wise mode of cleavage.

4.5.13 TTUDGA and TTUDGB - so similar, but so different uracil glycosylases of *T. thermophilus*

Despite that additional enzymatic tests and binding studies are required to understand the action of TTUDGA and, particularly, TTUDGB glycosylases, some conclusions can already be made at this stage of the work. As mentioned before, TTUDGA glycosylase is strictly uracil specific enzyme acting on double and single stranded DNA. High processivity, non-discriminative manner of action and mode of cleavage of uracil from different mispairs (4.5.10) makes this enzyme true functional homolog of Ung glycosylase. TTUDGB glycosylase is non-efficient, U/G mismatch specific enzyme, preferentially acting on double stranded DNA. These features make it similar to MUG/TDG enzymes. Thus, possibly, TTUDGA and TTUDGB in *T. thermophilus* form a couple of activities, similar way Ung and MUG in *E. coli*. Redundancy of uracil removal activities is widely distributed, but usually encoded by different, non-similar genes. *T. thermophilus* represents the microorganism, which contains two copies of highly similar genes, products of which, however, possess significant differences. Currently, the meaning of this is not known. Unfortunately, no additional information could be derived from the phylogenetic analysis (Fig. 29), because other microorganisms containing A and B versions of the enzyme (*C.*

crescentus, M. loti, P. aerophilum (unpublished data)) belong to different systematic groups, even kingdoms.

Certainly, number of highly intriguing questions can be raised already now. If one of the enzymes of *T. thermophilus* acts as a main defence line against deamination damages and another one serves as a back-up system, how do the organisms containing only "back-up" version counteracts uracil. Alternatively, one glycosylase could be freely moving activity and taking care of synthesised DNA molecules. Another version could be attached, let's say, to replication fork and participate in the damage removal during the DNA synthesis. In the light of fact that *T. thermophilus* contains high GC content and grows at high temperatures this scenario would be a reasonable possibility to avoid rapid cytosine deamination, especially in ssDNA. Here, it is worthy mentioning that this type of glycosylase was found as a part of replicative DNA polymerase I in bacteriophage SPO1 (Sandigursky and Franklin, 2000). Finally, the expression of particular enzyme version could correlate with certain life stages of the microorganism. As the work on TTUDGA and TTUDGB is only in its infancy, further research directed to clarify biological role of each glycosylase, possible interaction partners, catalytic mechanisms of action is undoubtfully highly advantageous and challenging.

4.5.14 Expression and purification of TTUDGA glycosylase in the ung$^-$ *E.coli* strain

4.5.14.1 Expression of the TTUDGA glycosylase in the BW 310 cells

Expression and purification of the TTUDGA protein (as well as TTUDGB) was performed in the BL21(DE3)pLysS strain (**4.5.4**). Unfortunately, as mentioned before, the strain is proficient in the ung gene. This is hardly a problem in case of TTUDGB enzyme, while its properties differs dramatic (**4.5.5**). Features of TTUDGA glycosylase, however, are quite similar to that of Ung enzyme. Thus, the alternative expression system was tested to avoid any possible contamination with the Ung activity. For this, the plasmid pET-21d/TTUDGA was transformed into several ung deficient strains. One of them, BW310 (Duncan and Weiss, 1982) was obtained from the **CGSC** (*E.coli* Genetic Stock Center) and is deficient only in the ung gene. Another one, CJ236, was donated by Dr.H.Kolmar and also lacks the dut (UTPase) gene (Joyce and Grindley, 1984).

As described before, expression of the target protein in the pET plasmid system is under the control of bacteriophage T7 promotor. Unfortunately, both new strains contain no sources of T7 RNA polymerase. Therefor, the plasmid pGP1-2 (obtained from Dr.S.Behrens) carrying T7 polymerase gene (Tabor and Richardson, 1985) was introduced into the both hosts (Fig. 40). The plasmid turned to be a suitable component due to the different antibiotic resistance gene and the different origin of replication from that of pET-21d vector.

An expression test was performed with both cells. Unfortunately, the transformants of CJ236 cells were growing only in case if pET-21d vector contained no insert. As soon as the pET-21d/TTUDGA construct was introduced, no growth of the culture was observed. Additional test revealed, that the cells lost their ability to grow in the presence of ampicillin, what indicates the lost of pET-21d/TTUDGA plasmid. The observation probably can be explained by the fact that dut deficient strain should contain large amounts of dUTP due to the lack of UTPase activity. The latter competes with TTP during the replication process (Kunkel *et.al.*, 1987) and, if incorporated, is not repaired by the appropriate glycosylase. It was described that every 20-30 nucleotide in the DNA of such hosts might be uracil residue (Duncan and Weiss, 1982). As the λ P_L promotor is known to be leaky (Tabor and Richardson, 1985), a certain amount of TTUDGA glycosylase is likely present in the cell. According to the observed high activity of the enzyme, numerous, highly cytotoxic AP sites could be produced. This, most probably, makes CJ263 cells not viable. The latter was not observed for the other strain. Thus BW310 was used for the following expression and purification experiments.

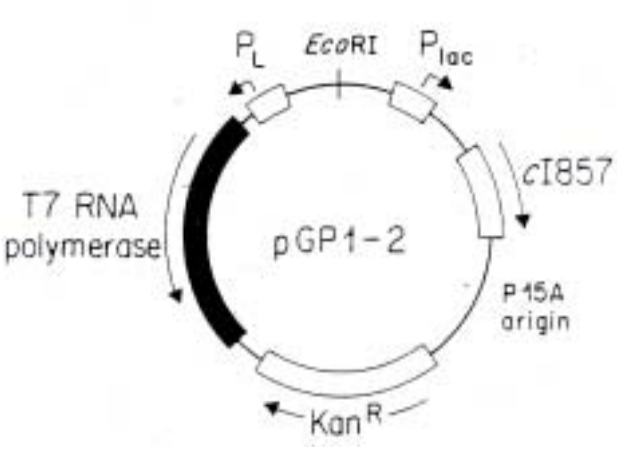


Fig. 40. Plasmid pGP1-2. Black box indicates the gene of T7 RNA polymerase, Kan^R - gene for aminoglycoside phosphotransferase, *c*I857 – heat-sensitive λ repressor gene, P_L – λ P_L promotor, P_{lac} – *lac* promotor. Taken from Tabor and Richardson, 1985.

In difference to the BL21(DE3)pLysS cells (Studier and Moffat, 1986), the gene of T7 RNA polymerase in the new expression system is placed under the temperature inducible P_L promotor (Tabor and Richardson, 1985). Consequently, the synthesis of the T7 RNA polymerase is induced by the increase of the growing temperature till 40-42 °C and denaturation of heat-sensitive λ repressor *c*I857 (Remaut *et.al.*, 1987). Then, highly specific T7 RNA polymerase transcribes the desired gene under the control of T7/*lac* promotor.

Accordingly, expression of TTUDGA glycosylase in the BW 310 cells was slightly modified. In short, the cells were incubated at 30 ^oC till OD_{600} reached 0.6. The protein synthesis was started with the increase of the temperatures till 40 ^oC and simultaneous addition of 1 mM IPTG. After 3 hours of incubation, the cells were harvested by the centrifugation and subjected to the following purification steps.

4.5.14.2 Purification of the TTUDGA glycosylase from BW 310 cells

The purification process of TTUDGA was similar to that described in the (**4.5.4**). The main difference appeared in the course of opening the cell walls. BW 310 strain contains no source of the T7 lysozyme. Thus, harvested cells were re-suspended in the working buffer and then, the lyses was started by the adding of the freshly prepared lysozyme solution (4 mg/ml). The mix was incubated for 15 min at the room temperature. The cells were further opened by sonication and the protein purified according the standard protocol. The purification scheme resulted in the active TTUDGA glycosylase which properties were undistinguishable from the earlier preparations (data not shown).

In general, pET-21d/TTUDGA/BW310 expression system is suitable to obtain the functional protein. However, the yield is very low (only about 0.1 mg of protein from the 1 l of the culture). Therefore, following optimisation steps are needed to improve the expression level of TTUDGA in the BW 310 cells.

4.6 Characterization of the surrounding genomic loci of genes encoding TTUDGA and TTUDGB glycosylases of *Thermus thermophilus*

Additional information about the biological role of glycosylases from *T. thermophilus* could be obtained by analysis surrounding genomic loci and its comparison with other organisms.

4.6.1 Surrounding of the gene encoding TTUDGA

At the 5' side of the gene of TTUDGA resides *orf* with unique sequence, which could code for a short protein of 120 a.a (for the sequence see Appendix). Possible functional interference with TTUDGA could be speculated from the fact that, stop codon of this *orf* is the part of start codon of TTUDGA gene with +1 shift in reading frame. Thus, both genes are transcribed into the same direction. In addition, similar sequence (63% identical and 80% similar residues) was found in closely related *D. radiodurans* genome and also in the close proximity to A type uracil - glycosylase. In case of *D. radiodurans*, the respective glycosylase and hypothetical *orf* are spaced by 89 bp, but also transcribed into the same direction. Thus, most likely, in both organisms glycosylase and putative protein with unknown function forms an operon. Possibly, this reflects adaptation requirements, specific for group, particularly, then no more representatives in more distant species were found to date.

As no clue about the possible function of these putative small proteins were obtained by homology searches, isolation and enzymatic characterisation of it is highly advantageous. In accordance with nomenclature of WIT/restricted database, the name X216 was introduced for the putative protein of *T. thermophilus*.

4.6.2 Cloning of the X216 gene from *T.thermophilus*

For the cloning primers were designed with *Nco*I and *Xho*I restriction sites at the 5' and 3' ends of the gene, respectively. PCR reaction was performed on the genomic DNA of *T.thermophilus* with DeepVent polymerase (NEB) and 5% DMSO. Resulting product was initially cloned into pCR-BluntII-Topo vector (Invitrogen). PCR screening revealed the presence of positive colonies, correctness of the insert was confirmed by DNA sequencing of the respective plasmids. Then, the gene X216 was re-cloned into the expression vector pET-21d.

4.6.3 Expression and purification of the X216

Expression and purification of the protein was done according standard methodology in BL21(DE3)pLysS cells (**Methods**). The only difference during the purification process was the use of thermo treatment (at 70 ^{o}C for 10 min) of crude lysate before loading on IMAC column. After that, X216 was purified till homogeneity onto a HS column (Fig. 41). Interesting, similar to TTUDGA, binding on the heparin column was inefficient. As usual, protein concentration and NaCl removal was done by Centriprep filter with cut – off value 3000. For long – time storage X216 was diluted with 50% glycerol and stored at –20 ^{o}C.

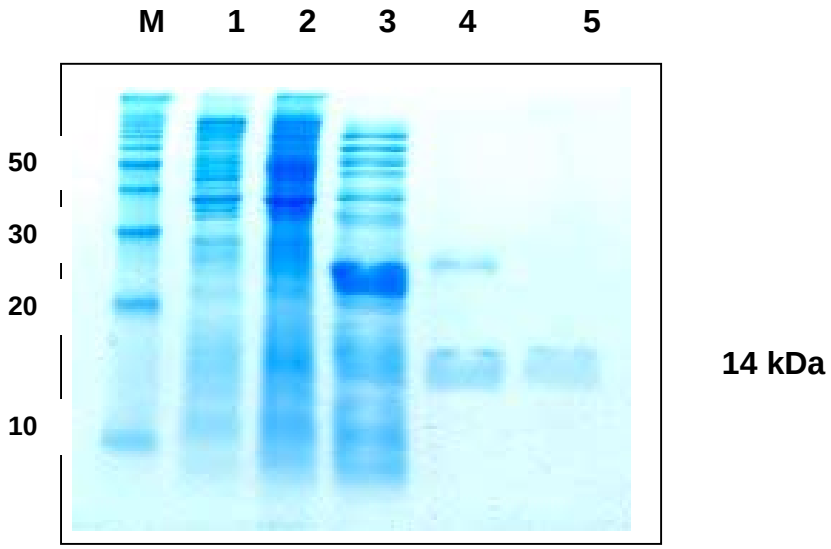


Fig. 41. Purification of X216 protein. **1**: crude extract of uninduced cells, **2**: crude extract of IPTG induced cells, **3**: supernatant after heat-treatment step, **4**: protein, eluted with 300 mM imidazole from IMAC, **5**: protein, eluted from HS column, **M**: 10 kDa protein size marker (Promega).

4.6.4 Test of glycosylase activity of X216

10 pmol of X216 protein was incubated with 40 fmol ss and ds DNA containing uracil under standart assay conditions **(Methods)**. No glycosylase activity was observed. Variations in reaction composition, like changes of buffer pH, temperature, addition of Mg^{+2} ions had no apparent influence.

4.6.5 Test of influence of X216 on the activity of TTUDGA glycosylase

1 pmol of TTUDGA glycosylase was mixed with 10 pmol X216 and incubated with two ds DNA substrates (U/G + U/A) at 50 ^{o}C, taking the samples at fixed time points within 10 min. The experiment was done in two different buffers with pH values 9.0 and 7.5 (room temp.). Peaks were integrated and cleavage efficiency of TTUDGA expressed graphically (Fig. 42).

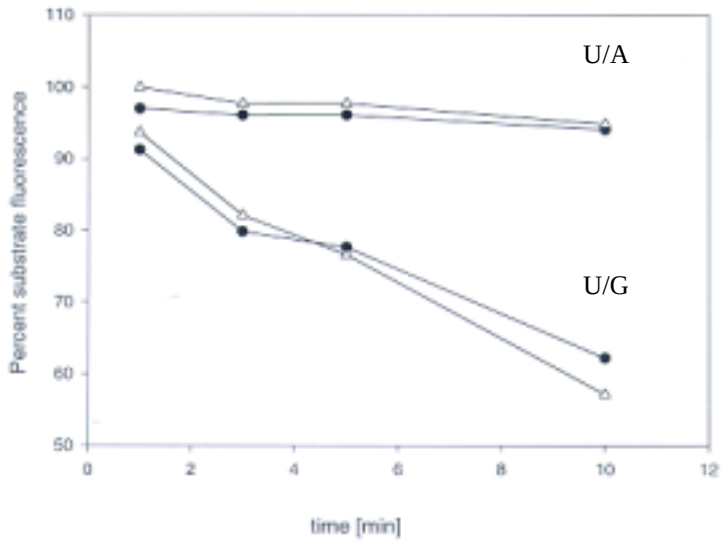


Fig. 42. Test of influence of X216 on the activity of TTUDGA glycosylase. Black circles indicate cleavage reaction without addition of X216, white triangles – with the addition of X216. Showed measurement performed in a buffer with pH 7.5. The data at pH 9.0 looks comparable (pH values are given at room temp.).

Experiment illustrated by Fig. 42 revealed that X216 protein brought no detectable difference in processing U/G and U/A mispairs by TTUDGA. However, it was only first test, similar experiment with ss DNA substrates would be highly advantageous.

In addition to above described experiments, activity of X216 on AP-site was tested. The initial results revealed intriguing possibility that the protein may act on AP-site in the ssDNA (data not shown). Unfortunately, the results were not always reproducible; therefore more experimental work is required for the unambiguous conclusion. Certainly, interaction studies of X216 protein with DNA and TTUDGA would be very interesting. Despite that highly efficient TTUDGA glycosylase hardly

requires additional stimulating factors, X216 could possibly serve as a scaffold for interaction of TTUDGA with other protein machinery.

4.6.6 Analysis of locus surrounding gene encoding TTUDGB

Annotation of the contig, containing gene of TTUDGB revealed the presence of several polymerase genes. Editing of the raw sequencing data resulted in joining of two *orfs* into one gene of DNA polymerase X (edited sequence is presented in Appendix). The latter belongs to the IV group of DNA polymerases, which comprises different nucleotidyl-transferases and eucaryotic polymerase β (Aravind and Koonin, 1998). Genes of TTUDGB and polymerase X are orientated in opposite direction, but makes 20 b.p. overlap with their 3' ends.

Polymerase X has quite limited distribution: it was found in *Thermus*, *Deinococcus*, *Bacillus*, *Aquifex* and *Methanobacterium* species (Aravind and Koonin, 1999). Unfortunately, no biochemical characterisation has been described. Polymerase X shears significant similarity to eucaryotic polymerase β, thus could be attributed as a possible DNA repair protein. Likely, it could be involved in the BER. Overlap with TTUDGB glycosylase in the genome of *T. thermophilus* speaks in favour to that. Alternatively, polymerase X and TTUDGB could function in different, yet unknown repair pathway. Finally, close genomic localisation could be accidental, particularly that is not conserved in other organisms. On the other hand, polymerase X stands in the close proximity to various DNA repair proteins in different species: *Methanosarcina mazei* – DNA ligase, *Methanobacterium thermoautotrophicum* – putative DNA helicase, *Bacillus* – MutS2. Thus, the role of polymerase X in the DNA repair processes - the questions presenting challenging area of research. Sequentially, deciphering the role of DNA polymerase X could possibly enable to understand probably complex biological role of TTUDGB protein.

5. Summary

Hydrolytic deamination reactions of cytosine and 5-methylcytosine results in highly mutagenic U/G and T/G mismatches, respectively. These processes are especially pronounced at high temperatures and present serious threats to genomic integrity for thermophiles. It is believed, that the respective damages are repaired by the base excision machinery (BER) which involves the action of uracil and T/G, U/G mismatches specific DNA glycosylases. With the aim to deepen knowledge about U/G and T/G removing enzymes, the focus of this work was done on isolation and enzymatic characterisation of several novel uracil glycosylases from two thermophilic microorganisms: archaeon *Methanobacterium thermoautotrophicum* ΔH and bacteria *Thermus thermophilus* HB27. *In vivo* role of characterised proteins was discussed by combining the enzymatic data with the comparative genomic analysis.

Part of the project, devoted to *M. thermoautotropicum* addressed the following main points:

Similarity search of known T/G, U/G specific enzymes on the completely sequenced genome of *M. thermoautotrophicum* (Smith *et.al*., 1997) revealed the presence of one candidate - *orf* 496. The latter was found to share high similarity to the helix-hairpin-helix class of U/G, T/G mismatches specific Mig glycosylases.

According to the sequence homology, *orf* 496 was re-annotated by the addition of 120 nucleotides to the 5' end, identification of putative rbs site and assignment of "*CTG*" as a new start codon.

Both truncated and full-length versions of the gene (named *mig*2) were cloned and overexpressed in *E. coli*. The originally annotated truncated version was found to be non-soluble, while the re-annotated *mig*2 expressed fully soluble protein, named Mig.*Mth*II. It was purified to apparent homogeneity.

It was shown that Mig.*Mth*II is a mono-functional U/G, T/G mismatch specific glycosylase. No significant activity on any other mismatch was found.

Influence of sequence context on the activity of Mig.*Mth*II was determined by multiple substrate kinetics. The assays revealed a highly discriminative nature of the

enzyme. Relative cleavage rates between the best and the worst DNA sequence patterns differed by a factor of 450 for the T/G, 80 - for U/G mismatch. Some of the substrates were not processed at all. Importantly, non-ability to act on particular DNA sequences is described for the first time among DNA repair enzymes. High requirements for the particular sequence context disabled to consider Mig.*Mth*II as a general U/G, T/G removing enzyme, but clearly indicated it's role in counteracting the damages, originating from cytosine methylation in the particular sequence context.

Additional confirmation of the assignment of biological role for Mig.*Mth*II was derived from re-annotation of *mig*2 gene. This resulted in a novel constellation, where the start codon of this gene serves as a stop codon of the putative cytosine specific methyltransferase (*orf* 495) with the frameshift +1. To test the functional interaction of Mig.*Mth*II and methylation events in the cell, the gene of the putative cytosine methyltransferase was isolated, cloned and expressed. The majority of the protein, named M.*Mth*H were not soluble. *In vivo* methylation assay, though, resulted in detectable methylation activity, but no consensus sequence was derived.

Comparative analysis of Mig glycosylases encoded by other microorganisms revealed that the latter are spread only among archaea; and always found in the close neighbourhood to cytosine specific methyltransferases. Some microorganisms (*M. mazei*, for example) carry identical co-localisation of glycosylase and cytosine specific methyltransferase genes as described in *M. thermoautotrophicum*. In such cases downstream placed *orfs* were also highly conserved and found to be putative RNA/DNA helicases. Their novel role in DNA repair or possible involvement in transcription/translation processes was discussed.

Part of the project, devoted to *T. thermophilus* addressed the following main point:

Two *orfs* with high similarity to recently described TMUDG uracil glycosylase (Sandigursky and Franklin, 1999) were identified in the genome of *T. thermophilus*. Both putative proteins are more related to the homologs from other microorganisms than in between, thus they clearly originate by different way than gene duplication.

Both *orfs* were cloned, overexpressed in *E. coli* and purified till homogeneity. Proteins were named putative TTUDGA and TTUDGB glycosylases.

Both enzymes were shown to be monofunctional uracil glycosylases, but TTUDGA was active on both ss and ds DNA, while TTUDGB performed efficiently only on ds DNA. It was estimated that TTUDGA acts about 800 times more efficient as TTUDGB.

Uracil removal efficiency from different oppositions was as following: in case of TTUDGA, the substrates ranked U/C>U/G>U/A, in case of TTUDGB, U/G>>U/C>U/A. Accordingly, it can be concluded that TTUDGA activity is primarily guided by the specificity to uracil and depends on the strength of hydrogen bonds inside the DNA helix. TTUDGB turned to be U/G mismatch specific glycosylase, thus action of the enzyme requires not only specific interaction with uracil, but also with the opposite base.

Importantly, TTUDGB lacks essential catalytic residues in the putative active centre. This fact in combination with the unique sequence composition and the enzymatic properties enabled to separate TTUDGB as a novel type of uracil glycosylase.

Analysis of the closest genomic surrounding of TTUDGA and TTUDGB in *T. thermophilus* was performed. In case of TTUDGA, direct physical overlap with *orf* X216 of unknown function was found. The latter has a homolog only in the genome of *D. radiodurans*, which is also located in the close proximity to the uracil glycosylase. Consequently, possible functional interaction between X216 and TTUDGA was suggested. X216 was cloned; the respective protein was expressed in *E. coli* and purified to homogeneity. Initial tests failed to detect glycosylase activity of X216 or it's influence on the activity of TTUDGA. However, the role of the X216 as a probable molecular matchmaker was suggested. Re-annotation and analysis of TTUDGB locus revealed co-localisation with DNA polymerase X. It’s high sequence similarity to human DNA polymerase β implies the possible involvement in DNA repair.

Orf with high similarity to Mig.*Mth*II glycosylase was identified in the genome of *T. thermophilus*. Gene cloning was followed by protein expression, purification and activity measurements. No activity on uracil or U/G, T/G mismatches was detected, however, A/G cleavage was measured. Accordingly, the protein was named MutY.*Tth* and, obviously, represents one of the defence mechanisms against damages caused by DNA oxidation in *T. thermophilus*.

6. Literature

Afeyan, N. B., Gordon, N. F., Mazsaroff, I., Varady, L., Fulton, S. P., Yang, Y. B. und Regnier, F. E. (1990). Flow-through particles for the high-performance liquid chromatographic separation of biomolecules: perfusion chromatography. *J Chromatogr.* **519**, 1-29.

Altschul St F., Madden T., Scheaffer A., Zhang J., Zhang Z., Miller W. and Lipman D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *NAR*, **27**(**17**), 3389 - 3402.

Aravind L. and Koonin E.V. (1999) DNA polymerase [beta]-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history, *NAR* **27**, 1609 - 1618.

Aravind L., Walker D.R. and Koonin E. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *NAR* **27**, 1223-1242.

Aravind L. and Koonin E.V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins, *NAR* **26**, 3746 - 3752.

Baranov P.V., Gurvich O.L., Fayet O., Prere M.F., Miller A., Gesterlan R.F., Atkins J.F. and Giddings M.C. (2001) RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *NAR* **29**, 264-267.

Barrett TE, Scharer OD, Savva R, Brown T, Jiricny J, Verdine GL, Pearl LH. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J*. **23**, 6599-6609.

Barrett T., Savva R., Panayotou G., Barlow T., Brown T., Jiricny J. and Pearl L. (1998) Crystal structure of a G:T/U mismatch - specific DNA glycosylase: mismatch recognition by complementary - strand interaction, *Cell* **92**, 117 - 129.

Behrens B, Noyer-Weidner M, Pawlek B, Lauster R, Balganesh TS, Trautner TA. (1987) Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. *EMBO J*. **6**, 1137-1142.

Benkovic S.J. and Cameron C.E. (1995) Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment *Escherichia coli* DNA polymerase I. *Meth.Enzymol*. **262**, 257-269.

Bernal A., Ear U. and Kyrpides N. (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *NAR* **29**, 126-127.

Bird L.E., Subramanya H.S. and Wigley D.B. (1998) Helicases: a unifying structural theme?, *Curr.Struc.Biol*. **8**, 14 - 18.

Bjelland S, Seeberg E. (1996) Different efficiencies of the Tag and AlkA DNA glycosylases from *Escherichia coli* in the removal of 3-methyladenine from single-stranded DNA. *FEBS Lett.* **11**, 127-129.

Bonfield J., Beal K. and Staden R. (1998) The new Staden Package manual.

Bond J.P. and Wallace S.S. (1999) Evolution of the Nth superfamily. ASM conference on DNA repair and mutagenesis. November 1-7, Hilton Head, South Caroline.

Breimer L.H., and Lindahl T. (1984) DNA glycosylase activities for thymine residues damaged by ring saturation, fragmentation, or ring contraction are functions of endonuclease III in *Escherichia coli*, *J. Biol. Chem.* **259**, 5543 – 5548.

Brennicke A., Marchfelder A. and Binder S. (1999) RNA editing, *FEMS Microbiology Reviews* **23**, 297 - 316.

Brock TD, Darland GK. (1970) Limits of microbial existence: temperature and pH. *Science* **169**, 1316-1318.

Brooks J.E., Blumenthal R.M. and Gingeras T.R. (1983) The isolation and characterization of the *Escherichia coli* DNA adenine methylase (*dam*) gene. *NAR* **11**, 837-851.

Brown J., Daniels Ch., and Reeve J. (1989) Gene structure, organization, and expression in archaebacteria, *Critical reviews in Microbiology* **16**, 287 – 337.

Carroll R. and Derse D. (1993) Translation of equine infectious anemia virus bicistronic tat mRNA requires leaky ribosome scanning of the tat CTG initiation codon. *J.Virol*. **67**, 1433-1440.

Cheng X. and Blumenthal R. (1999) S-adenosylmethionine-dependent methyltransferases: structures and functions. World Scientific Publishing Co.Pte.Ltd.

Cole,S.T., Eiglmeier,K., Parkhill,J., James,K.D., Thomson,N.R., Wheeler,P.R., Honore,N., Ganier,T., Churcher,C., Harris,D., Mungall,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R.M., Devlin,K., Duthoy,S., Feltwell,T., Fraser,A., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Lacroix,C., Maclean,J., Moule,S., Murphy,L., Oliver, Quail,M.A., Rajandream,M.-A., Rutherford,K.M., Rutter,S., Seeger,K., Simon,S., Simmonds,M., Skelton,J., Squares,R., Squares,S., Stevens,K., Taylor,K., Whitehead,S., Woodward,J.R. and Barrell,B.G. (2001) Massive gene decay in the leprosy *bacillus*. *Nature* **409**, 1007-1011.

Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E., Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrell,B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544.

Dar M., and Bhagwat A. (1993) Mechanism of expression of DNA repair gene *vsr*, an *Escherichia coli* gene that overlaps the DNA cytosine methylase gene, *dcm*, *Mol. Microbiology* **9**, 823 – 833.

David Sh. S. and Williams S. (1998) Chemistry of glycosylases and endonucleases involved in base - excision repair, *Chemical reviews* **98**, 1221 – 1261.

de la Cruz J, Kressler D, Linder P. (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *TiBS* **24**, 192-198.

de Laat W.L., Jaspers N.G. and Hoeijmakers J.H. (1999) Molecular mechanism of nucleotide excision repair. *Genes Dev* **13**, 768-785.

Delort A-M., Duplaa A-M., Molko D. and Teoule R. (1985) Excision of uracil residues in DNA: mechanism of action of *Escherichia coli* and *Micrococcus luteus* uracil-DNA glycosylase. *NAR* **13**, 319-335.

Dennis P.P. (1997) Ancient ciphers: translation in archaea. *Cell* **89**, 1007-1010.

Dieffenbach C.W. and Dveksler G.S. (1995) PCR primer. A laboratory manual. Cold Spring Harbor Laboratory Press.

Dodson, M. L., Michaels,M.L., and Lloyd,R.S. (1994) Unified catalytic mechanism for DNA glycosylases. *J.Biol.Chem*. **269,** 32709-32712.

Domena JD, Mosbaugh DW. (1985) Purification of nuclear and mitochondrial uracil-DNA glycosylase from rat liver. Identification of two distinct subcellular forms. *Biochemistry*. **25**, 7320-7328.

Doetsch P. and Cunningham R. (1990) The ezymology of apurinic/apyrimidinic endonucleases, *Mut. Research* **236**, 173 – 201.

Drotschmann K., Aronshtam A., Fritz H-J., Marinus M.G. (1998) The *Escherichia coli* MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *NAR* **26**, 948-953.

Duncan B.K. and Weiss B. (1982) Specific mutator effects of ung (uracil-DNA glycosylase) mutations in *Escherichia coli, J.Bacter*. **Aug**, 750 - 755.

Eftedal I., Guddal P.H., Slupphaug G., Volden G. and Krokan H. (1993) Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. *NAR* **21**, 2095-2101.

Eisen J.A. and Hanawalt P.C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes, *Mut.Research* **435**, 171 – 213.

Farooqui, A. A. (1980) Purification of enzymes by heparin-sepharose affinity chromatography. *J Chromatogr* **184**, 335-45.

Fondufe Yvonne. Characterization and direct modification of the substrate selectivity of Mig.*Mth*, a DNA repair glycosylase from the thermophilic archaeon *Methanobacterium thermoautotrophicum* THF, Dissertation zur Erlangung des Doktorgrades der Mathematisch – Naturwissenschaftlichen Fakultäten der Georg – August – Universität zu Göttingen, 1999, Göttingen.

Friedberg E., Walker G. And Siede W. (1995) DNA repair and mutagenesis. ASM Press, Washington.

Frostesjo L, Holm I, Grahn B, Page AW, Bestor TH, Heby O. (1997) Interference with DNA methyltransferase activity and genome methylation during F9 teratocarcinoma stem cell differentiation induced by polyamine depletion. *J Biol Chem*. **272**, 435943-66.

Gallinari P. and Jiricny J. (1996) A new class of uracil - DNA glycosylases related to human thymine - DNA glycosylase, *Nature* **383**, 735 - 738.

Gläsner W., Merkl R., Schellenberger V. and Fritz H.-J. (1995) Substrate preferences of Vsr DNA mismatch endonuclease and their consequences for the evolution of *the Escherichia coli* K-12 genome, *J. Mol. Biol*. **245**, 1 –7.

Gläsner W., Merkl R., Schmidt S., Cech D., Fritz H.J. (1992) Fast quantitative assay of sequence – specific endonuclease activity based on DNA sequncer technology. *Biol. Chem. Hoppe – Seyler* **373**, 1223 – 1225.

Goodman M.F. (1988) DNA replication fidelity: kinetics and thermodynamics. *Mutat. Res* **200**, 11-20.

Gogos A., Jantz D., Sentürker S., Richardson D., Dizdaroglu M. and Clarke N.D. (2000) Assignment of enzyme substrate specificity by principal component analysis of aligned protein sequences: an experimental test using DNA glycosylase homologs. *Proteins* **40**, 98-105.

Gold M. and Hurwitz J. (1964) The enzymatic methylation of ribonucleinic acid and deoxyribonucleic acid. Purification and properties of the deoxyribonucleic acid-methylating activity of *Escherischia coli*. *J.Biol.Chem*. **239**, 3858-3865.

Golinelli M.P., Chmiel N. and David Sh. (1999) Site – directed mutagenesis of the cysteine ligands to the [4Fe-4S] cluster of *Escherichia coli* MutY, *Biochemistry* **38**, 6997 – 7007.

Grigg G. and Clark S. (1994) Sequencing 5-methylcytosine residues in genomic DNA. *BioEssays* **16**, 431 - 436.

Grogan D.W. (1998) Hyperthermophiles and the problem of DNA instability, *MolMicrob*. **28**, 1043 - 1049.

Guan Y., Manuel R., Arvai A., Parikh S., Mol C., Miller J., Lloyd S., and Tainer J., (1998) MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily, *Nature Structural Biology* **5(12)**, 1058 – 1063.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166** (4), 557-80.

Hanck, T., Schmidt, S. und Fritz, H. J. (1993). Sequence-specific and mechanism-based crosslinking of Dcm DNA cytosine- C5 methyltransferase of *E. coli* K-12 to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine. *NAR* **21**, 303-309.

Haushalter K.A., Todd Stukenberg M.W., Kirschner M.W. and Verdine G.L., (1999) Identification of a new uracil – glycosylase family by expression cloning using synthetic inhibitors, *Curr.Biol*. **9**, p. 174-185.

Hendrich B., Hardeland U., Ng H-H., Jiricny J. and Bird A. (1999) The thymine glycosylase MBD4 can bind the product of deamination at methylated CpG sites. *Nature* **401**, 301-304.

Hennecke F., Kolmar H., Brundl K., and Fritz H.J. (1991) The vsr gene product of E. coli K - 12 is a strand - and sequence - specific DNA mismatch endonuclease, *Nature* **353**, 776 -778.

Hill J.W., Hazra T.K., Izumi T., and Mitra S. (2001) Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *NAR* **29**, 430-438.

Horst J.P. and Fritz H.J., (1996) Counteracting the mutagenic effect of hydrolytic deamination of DNA 5 - methylcytosine residues at high temperature: DNA mismatch N - glycosylase Mig.Mth of the thermophilic archaeon *Methanobacterium thermoautotrophicum* THF, *EMBO J.* 15, 5459 - 5469.

Horst J.P. Gentechnische Produktion und Characterisierung von Mig.Mth, einem Enzym zur Reparatur hydrolytischer Schäden an DNA 5 – Methylcytosinresten im thermophilen Archäon *Methanobacterium thermoautotrophicum* THF, Dissertation, 1996, Göttingen.

Hunter WN, Brown T, Kneale G, Anand NN, Rabinovich D, Kennard O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. *J Biol Chem*. **25**, 9962-9970.

Inouye, M., Arnheim, N., and Sternglanz,R. (1973) Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. *J.Biol.Chem*. **248,** 7247-7252.

Iost I. And Dreyfus M. (1994) mRNAs can be stabilized by DEAD-box proteins, *Nature* **372**, 193 – 196.

Jaenicke R. and Boehm G. (1998) The stability of proteins in extreme environments. *Curr.Opin.Struct.Biol*. **8**, 738-748.

Jiricny J. (1998) Replication errors: cha(lle)nging the genome. *EMBO J* **17**, 6427-6436.

Jones P.A., Rideout W.M., She J., Spruck C.H. and Tsai Y.C. (1992) Methylation, mutation and cancer. *Bioessays* **14**, 33-36.

Jones W.J., Nagle D.P. and Whitman W.B. (1987) Methanogens and the diversity of archaebacteria, *Microb. Reviews*, **Mar**., 135-177.

Joyce C.M. and Grindley N.D. (1984) Method for determining whether a gene of *Escherichia coli* is essential: application to the polA gene, *J.Bacter*. **May**, 636 – 643.

Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S., Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T., Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A., Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M., Takeuchi,C., Yamada,M. and Tabata,S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res*. **7**, 331-338.

Karran P. and Lindahl T. (1980) Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry* **19**, 6005-6011.

Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Kikuchi H, *et al*, (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix K1*., *DNA Res* **6**, 83 – 101, 145 – 152.

Klimasauskas S, Roberts RJ. (1995) M.*Hha*I binds tightly to substrates containing mismatches at the target base. *NAR* **25**, 1388-1395.

Klungland A., Höss M., Gunz D., Constantinou A., Clarkson S.G., Doetsch P.W., Bolton Ph.H., Wood R.D. and Lindahl T. (1999) Base excision repair of oxidative DNA damage activated by XPG protein, *MolCell* **3**, 33 – 42.

Krokan H.E., Standal R. and Slupphaug G. (1997) DNA glycosylases in the base excision repair of DNA, *Biochem J.* **325**, 1 – 16.

Kronberg A. (1992) DNA replication. W.H. Freeman and Co., San Francisko.

Kumar S., Cheng X., Klimasauskas S., Mi Sh., Posfai J., Roberts R., and Wilson G. (1994) The DNA (cytosine-5) methyltransferase. *NAR* **22**, 1-10.

Kumar S., Cheng X., Pflugrath J.W. and Roberts R.J. (1992) Purification, crystalization and preliminary X – ray diffraction analysis of an M.*Hha*I – AdoMet complex, *Biochemistry* **31**, 8648 – 8653.

Kunkel T. (1999) The high costs of living. *TIG* **15**, 93-94.

Kunkel T. (1992) DNA replication fidelity. *JBiolChem* **267**, 18251-18254.

Kunkel T., Roberts J. And Zakour R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meth.Enzym*. **154**, 367-382.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Laging Martin, (2000). Identifizierung und Characterisierung eines Vsr-Homologen aus *Bacillus stearothermophilus*. Dissertation zur Erlangung des Doktorgrades der Mathermatisch-Naturwissenschaftlichen Fakultäten der Georg-August-Universität zu Göttingen.

Lathe III, W.C., Snel B. and Bork P. (2000) Gene context conservation of a higher order than operons, *TIBS* **25**, 474 – 479.

Lauster R, Trautner TA, Noyer-Weidner M. (1989) Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *JMolBiol*. **20**, 305-312.

Lieb M. and Bhagwat A. (1996) Very short patch repair: reducing the costs of cytosine methylation. *MolMicrob* **20**, 467-473.

Lieb M, Allen E, Read D. (1986) Very short patch mismatch repair in phage lambda: repair sites and length of repair tracts. *Genetics* **114**, 1041-1060.

Lindahl T. and Wood R. (1999) Quality control by DNA repair. *Science* **286**, 1897-1905.

Lindahl T. (1993) Instability and decay of primary structure of DNA, *Nature* **362**, 709 - 715.

Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens B. (1977) DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J.BiolChem*. **252**, 3286-3294.

Lindahl T. (1974) An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues, *PNAS* **71**, 3649 - 3653.

Lindahl T. and Nyberg B. (1974) Heat induced deamination of cytosine residues in deoxyribonucleic acid, *Biochemistry* **13**, 3405 - 3410.

Lindahl T, Nyberg B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **12**, 3610-3618.

Loeb LA, Kunkel TA. (1982) Fidelity of DNA synthesis. *Annu Rev Biochem*. **51**, 429-57.

Lustig K.D., Stukenberg P.T., McGarry T.J., King R.W., Cryns V.L., Mead P.E. (1997) Small pool expression screening: identification of genes involved in cell cycle control, apoptosis, and early development. *Meth.Enzymol*. **283**, 83-99.

Lyko F., Ramsahoye B.H. and Jaenisch R. (2000) DNA methylation in *Drosophila melanogaster*. *Nature* **408**, 538-539.

Malone T., Blumenthal R and Cheng X. (1995) Structure - guided analysis reveals nine sequence motifs conserved among DNA amino - methyl - transferases, and suggest a catalytic mechanism for these enzymes, *J Mol Biol* **253**, 618 - 632.

Marnet L.J. and Plastaras J.P. (2001) Endogenous DNA damage and mutation. *TIG* **17**, 214-221.

Michaels M.L., Tchou J., Grollman A.P., and Miller J.H. (1992) A repair sytem for 8 – oxo – 7,8 – dihydrodeoxyguanine, *Biochemistry* **31**, 10964 – 10968.

Modrich P. and Lahue R. (1996) Mismatch repair in replication fidelity, genetic recombination and cancer biology. *Annu Rev Biochem* **65**, 101-133.

Mol C.D., Arval A.S., Slupphaug G., Kavli B., Alseth I., Krokan H. and Tainer J. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869-878.

Moore MH, Gulbis JM, Dodson EJ, Demple B, Moody PC. (1994) Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanine-DNA methyltransferase from *E. coli*. *EMBO J.* **13**, 1495-1501.

Mullis, K. B. und Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155,** 335-50.

Nash H.M., Bruner S.D., Schärer O.D., Kawate T., Addona T.A., Spooner E., Lane W.S., Verdine G. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr.Biol* **6**, 968-980.

Neddermann P, Gallinari P, Lettieri T, Schmid D, Truong O, Hsuan JJ, Wiebauer K, Jiricny J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J.Biol.Chem.* **271**, 12767-12774.

Nelson and Cox (2000) Lehninger principals of biochemistry.

Niederreither K., Harbers M., Chambon P., and Dolle P. (1998) Expression of T:G mismatch – specific thymidine – DNA glycosylase and DNA methyl transferase genes during development and tumorigenesis, *Oncogene* **17**, 1577 – 1585.

Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,Nd, Ely,B., Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and Fraser,C.M. (2001) Complete Genome Sequence of *Caulobacter crescentus. PCNA* U.S.A. **99** (in press).

Nilsen H., Yazdankhah S., Eftedal I., Krokan H. (1995) Sequence specificity for removal of uracil from U/A pairs and U/G mismatches by uracil – DNA glycosylase

from *Escherichia coli*, and correlation with mutational hotspots, *FEBS Letters* **362**, 205 – 209.

Noll D.M., Gogos A., Granek J.A. and Clarke N.D. (1999) The C-terminal domain of the adenine-DNA glycosylase MutY confers specificity for 8-oxoguanine-adenine mispairs and may have evolved from MutY, an 8-oxo-dGTPase. *Biochemistry* **38**, 6374-6379.

Nölling J., Eeden F., Eggen R. and Vos W. (1992) Modular organisation of related archaeal plasmids encoding different restriction - modification systems in *Methanobacterium thermoformicicum*, *NAR* **20**, 6501 - 6507.

Nölling J., and de Vos W., (1992) Characterization of the archael, plasmid – encoded type II restriction – modification system *MTHT*I from *Methanobacterium thermoformicicum* THF: homology to the bacterial *NgoPII* system from *Neisseria gonorrhoeae, J: Bacteriology* **Sept**, 5719 – 5726.

Nölling,J., Frijlink,M., and Vos,W.M. (1991) Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. *J. Gen. Microbiol*., **137**, 1981-1986.

Olsen G.J. and Woese C.R. (1997) Archaeal genomics: an overview. *Cell* **89**, 991-994.

Pace C.N., Vajdos F., Fee L., Grimsley G., Gray T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci*. **4**, 2411 – 2423.

Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

Panayotou G, Brown T., Barlow T., Pearl L.H. and Savva R. (1998) direct measurement of the substrate preferences of uracil-DNA glycosylase. *J.Biol.Chem*. **273**, 45-50.

Parikh S.S., Mol C.D., Hosfield D.J. and Tainer J.A. (1999) Envisioning the molecular choreography of DNA base excision repair, *Curr.Struc.Biol*. **9**, 37 – 47.

Pearl L.H., (2000) Structure and function in the uracil-DNA glycosylase superfamily, *MutResearch* **460**, 165 – 181.

Pfennig,G.N. and Lippert,K.D. (1966) Über das Vitamin B_{12}-Bedürfnis phototropher Schwefelbakterien. *Arch. Microbiol*., **55**, 245-256.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599.

Raleigh E.A. (1998) Why are there so many restriction emdonucleases. *The NEB Transcript* **9**, 6-7.

Rein Th., DePamphilis M.L. and Zorbas H. (1998) Identifying 5-methylcytosine and related modifications in DNA genomes. *NAR* **26**, 2255-2264.

Remaut E.,.Stanssens P. and Fiers W. (1981) Plasmid vectors for high-efficiency expression controlled by the P_L promotor of coliphage lambda. *Gene* **15**, 81-93.

Roberts J. and Macelis D. (2001) REBASE – restriction enzymes and methylases. *NAR* **29**, 268-269.

Roberts J. (1995) On base flipping. *Cell* **82**, 9-12.

Russel, M. and Model, P. (1984) Replacement of the *fip* gene of *Escherichia coli* by an inactive gene cloned on a plasmid. *J. Bacteriol*. **159**, 1034-1039.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory Press, NY.

Sandigursky M. and Franklin W. (2000) Uracil-DNA glycosylase in the extreme thermophile *Archaeoglobus fulgidus*. *J.Biol.Chem* **275**, 19146-19149.

Sandigursky M., and Franklin W. (1999) Thermostable uracil – DNA glycosylase from *Thermotoga maritima*, a member of a novel class of DNA repair enzymes, *Current Biology* **9**, 531 – 534.

Saparbaev M. and Laval J. (1998) 3,N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *Proc Natl Acad Sci U S A.* **21**, 8508-8513.

Savva, R., McAuley-Hecht, K., Brown, T. and Pearl, L. (1995) The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487-493.

Schellenberger V., Siegel R.A., Rutter W.J., (1993) Analysis of enzyme specificity by multiple substrate kinetics, *Biochemistry* **32**, 4344 – 4348.

Schärer O.D. and Jiricny J. (2001) recent progress in the biology, chemistry and structural biology of DNA glycosylases. *BioEssays* **23**, 270-281.

Shen J.Ch., Rideout W.M. and Jones P.A. (1994) The rate of hydrolytic deamination of 5-methylcytosien in double-stranded DNA. *NAR* **22**, 972-976.

Shibutani S, Takeshita M, Grollman AP. (1991) Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **31**, 431-434.

Shinohara A. and Ogawa T. (1995) Homologous recombination and the roles of double-strand breaks. *TIBS* **20**, 387-391.

Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E. and Tainer, J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87-92.

Smith D., Doucette-Stamm L., Deloughery C., Lee H., Dubois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K., Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D., Spadafora R., Vicaire R., Wang Y., Wierzbowski J., Gibson R., Jiwani N., Caruso A., Bush D., Reeve J., (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics, *J. Bacteriology* **179**, 7135 - 7155.

Snel B., Bork P. and Huynen M (2000) Genome evolution: gene fusion versus gene fission. *TIG* **16**, 9-11.

Sohail A., Lieb M., Dar M., and Bhagwat A. (1990) A gene required for very short patch repair in *E. coli* is adjacent to the DNA cytosine methylase gene, *J. Bacteriology* **172**, 4214 – 4221.

Studier,F.W., Rosenberg,A.H., Dunn,J.J., and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol*. **185**, 60-89.

Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189** (1), 113-30.

Sweeny, W.V. (1980) Proteins containing 4Fe-4S clusters: an overview. *Ann. Rev. Biochem.* **49**, 139-161.

Tabor S. and Richardson Ch.C. (1985) A bacteriophage T7 RNA polymerase/promotor system for controlled exclusive expression of specific genes, *PNAS* **82**, 1074 – 1078.

Tate W.P., Mansell J.B., Mannering S.A., Irvine J.H., Major L.L. and Wilson D.N. (1999) UGA: a dual signal for “stop” and for recoding in protein synthesis. *Biochemistry* **64**, 1342-1453.

Tatiana A. Tatusova, Thomas L. Madden. (1999) Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*. **174,** 247-250.

Thayer M., Ahern H., Xing D., Cunningham R. and Tainer J. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure, *EMBO J*. **14**, 4108 – 4120.

Thompson,J.D., Gouy M., Higgins,D.G., Gibson,T.J. (1998) Multiple sequence alignment with Clustal X. *TIBS* **23**, 403-405.

Thompson,J.D., Higgins,D.G., Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *NAR* **22**, 4673-4680.

Tsutakawa S.E., Muto T., Kawate T., Jingami H., Kunishima N., Ariyoshi M., Kohda D., Nakagawa M., and Morikawa K. (1999) Crystallographic and functional studies of Very Short Patch Repair endonuclease. *Mol.Cell* **3**, 621-628.

Tye BK., Chien J., Lehman I.R., Duncan B.K. and Warner H.R. (1978) Uracil incorporation: a source of pulse-labeled DNA fragments in the replication of the *Escherichia coli* chromosome. *Proc.Natl.Acad.Sci.USA* **75**, 233-237.

Tye BK, Lehman IR. (1977) Excision repair of uracil incorporated in DNA as a result of a defect in dUTPase. *J Mol Biol*., **117**, 293-306.

Ullah S., Gallinari P., Xu Y.-Z., Goodman M., Bloom L. Jiricny J., and Day R., (1996) Base analog and neighboring base effects on substrate specificity of recombinant human G:T mismatch – specific thymine DNA – glycosylase, *Biochemistry* **35**, 12926 – 12932.

Vanyushin BF, Kirnos MD. (1988) DNA methylation in plants. *Gene* **25**, 117-211.

Vanyushin BF, Belozersky AN, Kokurina NA, Kadirova DX. (1968) 5-methylcytosine and 6-methylamino-purine in bacterial DNA. *Nature* **218**, 1066-1067.

Waters T.R. and Swann P.F. (2000) Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *MutResearch* **462**, 137-147.

Wetlaufer D.B. (1962) Ultraviolet spectra of proteins and amino acids. *Adv.Protein.Chem.* **17**, 303-391.

Williams S.D. and David Sh.S. (1998) Evidence that MutY is a monofunctional glycosylase capable of forming a covalent Schiff base intermediate with substrate DNA, *NAR* **26**, 5123 – 5133.

White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin H, Jiang L, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Fraser CM.,

(1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1, *Science* **286,** 1571-7.

Wolin,E.A., Wolfe,R.R., and Wolin,M.J. (1964) Viologen dye inhibition of methane formation by *Methanobacterium omelanskii*. *J. Bacteriol*., **87**, 993-998.

Yang H., Clendenin WM., Wong D., Demple B., Slupska M.M., Chiang J.H. and Miller J.H. (2001) Enhanced activity of adenine-DNA glycosylase (Myh) by apurinic/apyrimidinic endinuclease (Ape1) in mammalian base excision repair of an A/GO mismatch. *NAR* **29**, 743-752.

Yang H., Fitz – Gibbon S., Marcotte E.M., Tai J.H., Hyman E.C., and Miller J.H. (2000) Characterization of a thermostable DNA glycosylase specific for U/G and T/G mismatches from hyperthermophilic archaeon *Pyrobaculum aerophilum, J. Bacteriology* **Mar**., 1272 – 1279**.**

Zeikus, J.G., and Wolfe, R. S. (1972) *Methanobacterium autotrophicus sp*. n., an anaerobic, autotrophic, extreme thermophile. *J. Bacteriol*. **109**, 707-713.

7. Appendix

7.1 Abbreviations

A	Adenine
A.L.F.	Automated Laser Fluorescent Sequencer
AMP	β-Lactamase gene
APS	Ammoniumperoxodisulphate
ATP	Adenosine-5'-phosphate
BCIP	5-Brom-4-chlor-3-indolyl-phosphate
bp	Base pair
C	Cytosine
°C	degree Celsius
CV	column volume
Dcm	DNA-Cytosine-Methyltransferase
DMSO	Dimethylsulfoxide
DNA	Desoxyribonucleic acid
dNTP's	2'-Desoxyribonucleoside-5'-trisphosphate
dTTP	2'-Desoxytymidine
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylendiamintetraacetate
e.g	for example
e-value	similarity by chance
f	femto (10^{-15})
*f*1 *ori*	replication origin of bacteriophage f1
Fig	Figure
FITC	Fluorescein
g	Gramme
G	Guanine
h	hour
HEPES	4-(2-Hydroxyethyl)-1-piperazine-ethane-sulphonic acid
IMAC	Immobilized Metall Affinity Chromatography

i.e.	that is
IPTG	Isopropyl-β-thiogalactopyranoside
kb	Kilobase pair
l	Liter
lac	β-Galactosidase gene
M	molar
4-meC	4-N-methylcytosine
5-meC	5-C-methylcytosine
µ	micro (10^{-6})
m	mili (10^{-6})
min	minute
ml	Milliliter
M_r	relative molecular weight
n	nano (10^{-9})
NBT	nitroblue-tetrazolium
OD_X	optical density at a particular wavelength X nm
ori	origin of replication
p	pico (10^{-12})
PAGE	polyacrylamide gel elektrophorese
PCR	Polymerase Chain Reaction
RNA	ribonuclein acid
s	second
SAM	S-adenosylmethionine
SDS	sodium dodecylsulphate
ssDNA	single stranded-DNA
T	Thymine
t	time
Tab	Table
TBE	Tris/boric acid buffer
TE	Tris/EDTA-buffer
TEMED	N, N, N', N'-Tetramethylethylendiamine
Tris	Tris-hydroxymethyl-aminomethane
u	*units*

U	uracil
UV	ultraviolet light
V	volt
X-gal	5-Brom-4-Chlor-3-Indolyl-β-D-Thiogalactopyranoside
8-oxo-G	7,8-dihydro-8-Oxoguanine

7.2 Program for protein purification on Biocad™ Workstation

General settings: set detectors inline
General settings: set column offline
General settings: set SCOUT column 1A: Poros (**name and diameter of column**[1])
General settings: set column 1 inline
General settings: UV detector wavelength = 280 nM and 260 nM
General settings: flow rate = 10.00 ml/min
General settings: turn UV detector lamp on
Pump on

Method Start
[01] Equilibration Block
(A) Equilibration with the buffer[2]
0.0 CV[3] Set solvent blend 25% 100 mM buffer, 75% water
0.0 CV Set detectors inline
4.0 CV Zero UV detectors
4.0 CV End solvent blend 25% 100 mM buffer, 75% water

[02] Load Block
(A) Step segment
0.0 ml Load through pump 100% sample
0.0 Start fraction collection, each 8.0 ml
(**volume of sample**) ml Load through pump 100% sample

[03] Wash Block
(A) Wash with buffer
0.0 CV Set solvent blend 25% 100 mM buffer, 75% water
10.0 CV end solvent blend 25% 100 mM buffer, 75% water

[04] Elute Block
(A) Gradient segment
0.0 CV Start gradient 25% 100 mM buffer, 75% water
0.0 CV Start fraction collection peak (**volume of fraction**) ml Thr (**absorption value at 280 nm**) slope[4]
15.0 CV End gradient 25% 100 mM buffer, 25% water, 50% 3M NaCl

(B) Wash

0.0 CV Set solvent blend 25% 100 mM buffer, 5% water, 70% 3M NaCl

5.0 CV Stop fraction collection

5.0 CV end solvent blend 25% 100 mM buffer, 5% water, 70% 3M NaCl

(C) Wash with buffer

0.0 Set solvent blend 25% 100 mM buffer, 75% water

5.0 End solvent blend 25% 100 mM buffer, 75% water

[1] parameters, which changes according the type of protein purified and column used. Indicated in bold letters

[2] in this work 25 mM HEPES, pH 7.6.

[3] column volume.

[4] at desired values of absorbtion at 280 nm, the fraction volume can be reduced.

7.3 Principle of the multiple substrate kinetics

Specificity of the enzyme can be studied by the means of multiple substrate kinetics and expressed as the relative values. The principle of the method allows to compare the reactivity of several substrates in one reaction (Schellenberger *et.al.*, 1993). Normally, the consumption of substrates in the course of incubation can be expressed by the following equation:

$$\frac{d[S_i]}{dt} = -k_i \cdot [E] \cdot [S_i] \qquad \textbf{[1]}$$ where:

$[S_i]$ = Concentration of substrate i
$[E]$ = Concentration of free enzyme
i = 1...n

$k_i = \frac{k_{cat}}{K_M}$ second-order rate constant for the reaction between substrate and free enzyme.

During the reaction, concentrations of substrates are changing. The amounts of free and bound enzyme concentrations are also changing and may be impossible to calculate. Thus, determination of an absolute k_i value for the each substrate (namely, the specificity and efficiency of the enzyme to the particular substrate) can be also impossible. However, comparison among substrates can be done by the ratios of their second-order constants. Equation [1] gives the rate of consumption of two simultaneously processed substrates:

$$d[S_1]/d[S_2] = k_1[S_1]/k_2[S_2] \qquad \textbf{[2]}$$

Separate integration of each part of the equation [2] leads to:

$$\int_0^t \frac{d[S_1]}{[S_1]} = -\int_0^t k_1 \cdot E \cdot dt$$

$$\ln \frac{[S_1]_t}{[S_1]_0} = -k_1 \cdot E \cdot t$$

$$-E \cdot t = \frac{1}{k_1} \cdot \ln \frac{[S_1]_0}{[S_1]_t}$$

For both parts of equation [2]:

$$\frac{1}{k_1} \cdot \ln \frac{[S_1]_t}{[S_1]_0} = \frac{1}{k_2} \cdot \ln \frac{[S_2]_t}{[S_2]_0} \qquad \textbf{[3]}$$

Then, $k_{rel} = k_1/k_2$ and can be obtained by measuring the concentrations of both substrates at time points 0 and t. If k_1 represents the reference substrate and is normalised to be k_1=1, then comparison among several substrates can be done and measurements obtained from different reactions merged together.

7.4 Nucleotide sequences of loci of DNA glycosylases genes

Positions of the respective primers used for PCR and DNA sequencing reactions (see **Materials**) are indicated by bold letters. Below nucleotide sequence, the protein sequences are given.

7.4.1 Nucleotide sequence of genes encoding Mig.*Mth*II DNA glycosylase and M.*Mth*H DNA methyltransferase of *Methanobacterium thermoautotrophicum* ΔH

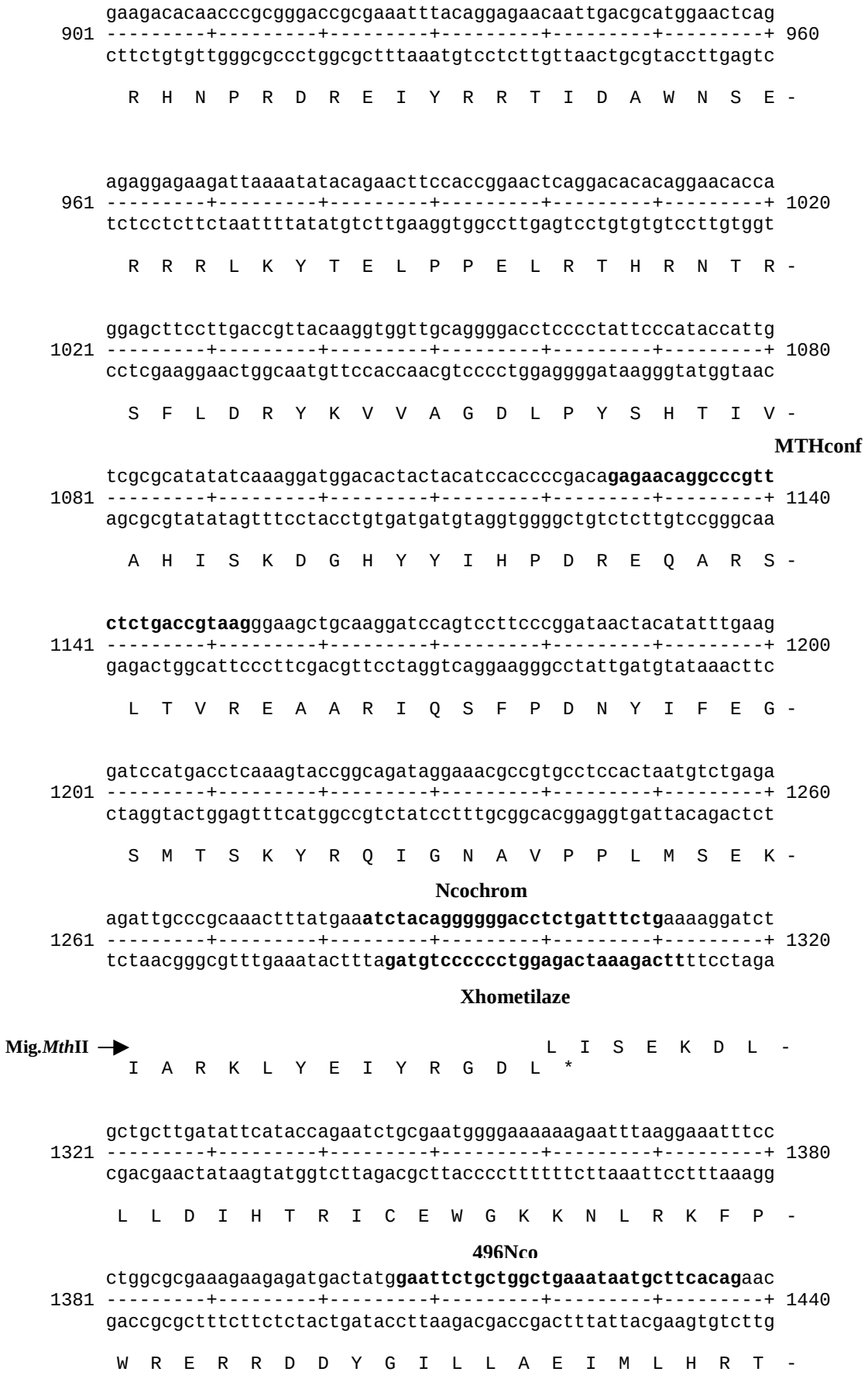
Ncometilaz

```
   aaaataatttaatgtggttaagataaatcatcaatgacaggtaacca gaggcaggttcgg
 1 ---------+---------+---------+---------+---------+---------+ 60
   ttttattaaattacaccaattctatttagtagttactgtccattggtctccgtccaagcc

   cgatgtcagatgttgtg ttgattgacatatttgcaggtgcgggggggcttactgaaggtt
61 ---------+---------+---------+---------+---------+---------+ 120
   gctacagtctacaacacaactaactgtataaacgtccacgccccccgaatgacttccaa

     M  S  D  V  V  L  I  D  I  F  A  G  A  G  G  L  T  E  G  F -
```

M.*Mth*H →

```
   tccttcggagtgattatacattcgtttcacatattgagatggacagggatgcgattcaga
121 ---------+---------+---------+---------+---------+---------+ 180
   aggaagcctcactaatatgtaagcaaagtgtataactctacctgtccctacgctaagtct

     L  R  S  D  Y  T  F  V  S  H  I  E  M  D  R  D  A  I  Q  T -

   cactggagacaaggggactctatcatcacctgcagagcgatggtgaccccgaggattaca
181 ---------+---------+---------+---------+---------+---------+ 240
   gtgacctctgttcccctgagatagtagtggacgtctcgctaccactggggctcctaatgt

     L  E  T  R  G  L  Y  H  H  L  Q  S  D  G  D  P  E  D  Y  T -

   ccgaatacgttaatggtgagataggacgtgaagaactattcagaagatatcctgactttg
241 ---------+---------+---------+---------+---------+---------+ 300
   ggcttatgcaattaccactctatcctgcacttcttgataagtcttctataggactgaaac

     E  Y  V  N  G  E  I  G  R  E  E  L  F  R  R  Y  P  D  F  D -

   acagtgaactgtacatgaaccttgaactcacagaggaaaacgtggacagggtcatagaga
301 ---------+---------+---------+---------+---------+---------+ 360
   tgtcacttgacatgtacttggaacttgagtgtctcctttgcacctgtcccagtatctct

     S  E  L  Y  M  N  L  E  L  T  E  E  N  V  D  R  V  I  E  T -
```

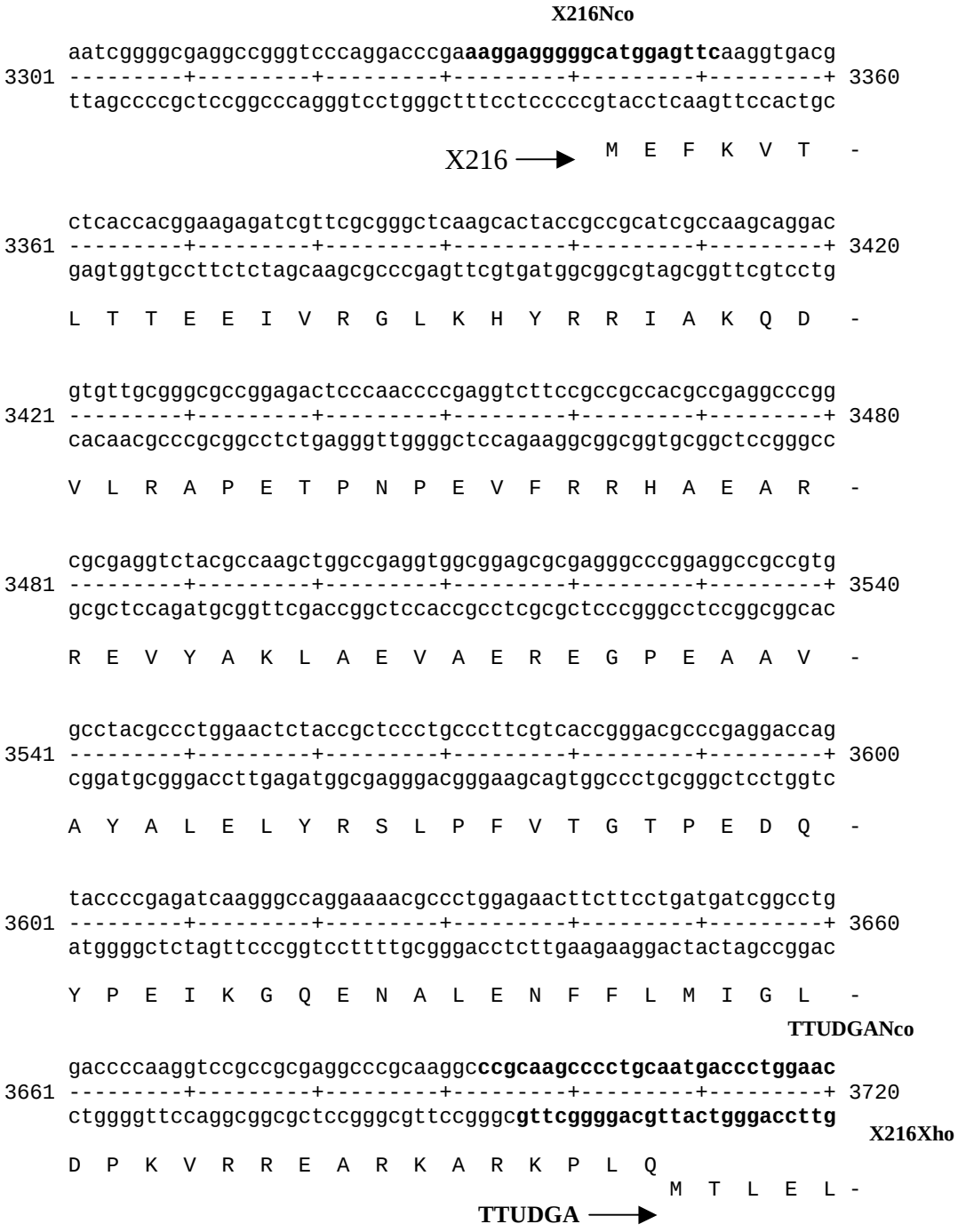
```
    ccatcaggtcaaagatgaatgatatggggacagtttctgttgatggcataataggtgggc
361 ---------+---------+---------+---------+---------+---------+ 420
    ggtagtccagtttctacttactatacccctgtcaaagacaactaccgtattatccacccg

      I  R  S  K  M  N  D  M  G  T  V  S  V  D  G  I  I  G  G  P -


    ctccatgtcaggcgtactcatatgctggcagatcccggaagaacatggaaaaggacagga
421 ---------+---------+---------+---------+---------+---------+ 480
    gaggtacagtccgcatgagtatacgaccgtctagggccttcttgtacctttcctgtcct

      P  C  Q  A  Y  S  Y  A  G  R  S  R  K  N  M  E  K  D  R  R -


    gaaactacctctacctcctgtacattaaattcctgaaagaattccagccggaattctttg
481 ---------+---------+---------+---------+---------+---------+ 540
    ctttgatggagatggaggacatgtaatttaaggactttcttaaggtcggccttaagaaac

      N  Y  L  Y  L  L  Y  I  K  F  L  K  E  F  Q  P  E  F  F  V -


    tattcgagaacgtccccgggatgaagtcagcgcaaaggggccacatactatctgatttcc
541 ---------+---------+---------+---------+---------+---------+ 600
    ataagctcttgcaggggccctacttcagtcgcgtttccccggtgtatgatagactaaagg

      F  E  N  V  P  G  M  K  S  A  Q  R  G  H  I  L  S  D  F  Q -


    agaggaaggtaacagacctggaatataaacttgactttgaagtgcgtgatgcatataact
601 ---------+---------+---------+---------+---------+---------+ 660
    tctccttccattgtctggaccttatatttgaactgaaacttcacgcactacgtatattga

      R  K  V  T  D  L  E  Y  K  L  D  F  E  V  R  D  A  Y  N  F -


    tcggggtccctcagaggagaaggaggatcatagtcatcggccaccgcatggaagataaaa
661 ---------+---------+---------+---------+---------+---------+ 720
    agccccagggagtctcctcttcctcctagtatcagtagccggtggcgtaccttctatttt

      G  V  P  Q  R  R  R  R  I  I  V  I  G  H  R  M  E  D  K  R -


    ggatacagtttgatgaggagaggtatgccggtacagtgaatgacatcctatgtgatctgc
721 ---------+---------+---------+---------+---------+---------+ 780
    cctatgtcaaactactcctctccatacggccatgtcacttactgtaggatacactagacg

      I  Q  F  D  E  E  R  Y  A  G  T  V  N  D  I  L  C  D  L  P -


    ccgcccttgaacccggccagggcacagacgggccccaggaatatagatccagaccatcac
781 ---------+---------+---------+---------+---------+---------+ 840
    ggcgggaacttgggccggtcccgtgtctgccccggggtccttatatctaggtctggtagtg

      A  L  E  P  G  Q  G  T  D  G  P  Q  E  Y  R  S  R  P  S  R -


    gcctcctgaatgagatgggtataaggacagagaatgatattctccgacatcaccaggcgc
841 ---------+---------+---------+---------+---------+---------+ 900
    cggaggacttactctacccatattcctgtctcttactataagaggctgtagtggtccgcg

      L  L  N  E  M  G  I  R  T  E  N  D  I  L  R  H  H  Q  A  R -
```

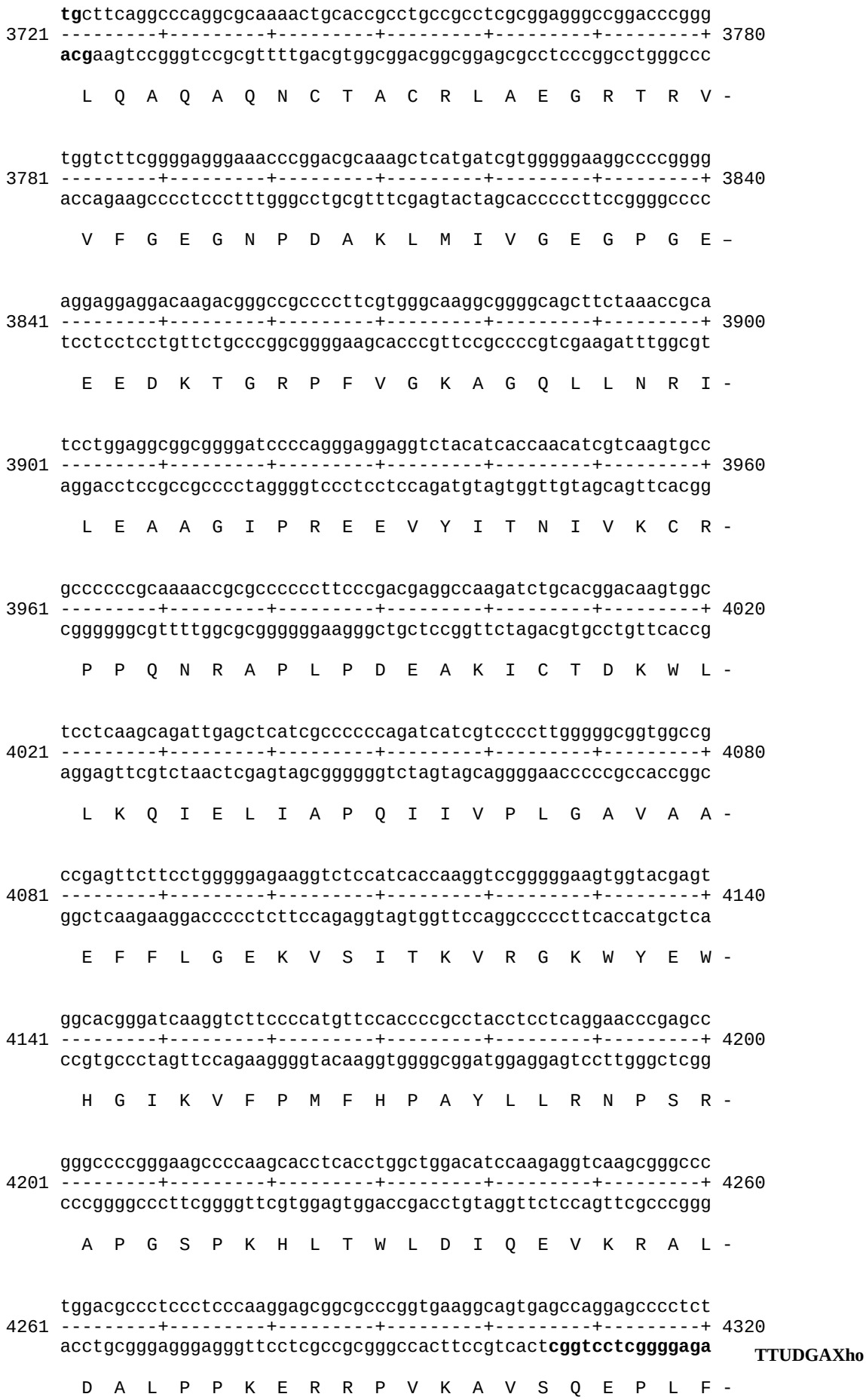

gaagacacaacccgcgggaccgcgaaatttacaggagaacaattgacgcatggaactcag
901 ---------+---------+---------+---------+---------+---------+ 960
cttctgtgttgggcgccctggcgctttaaatgtcctcttgttaactgcgtaccttgagtc
R H N P R D R E I Y R R T I D A W N S E -
agaggagaagattaaaatatacagaacttccaccggaactcaggacacacaggaacacca
961 ---------+---------+---------+---------+---------+---------+ 1020
tctcctcttctaattttatatgtcttgaaggtggccttgagtcctgtgtgtccttgtggt
R R R L K Y T E L P P E L R T H R N T R -
ggagcttccttgaccgttacaaggtggttgcaggggacctcccctattcccataccattg
1021 ---------+---------+---------+---------+---------+---------+ 1080
cctcgaaggaactggcaatgttccaccaacgtcccctggaggggataagggtatggtaac
S F L D R Y K V V A G D L P Y S H T I V -
MTHconf
tcgcgcatatatcaaaggatggacactactacatccaccccgacagagaacaggcccgtt
1081 ---------+---------+---------+---------+---------+---------+ 1140
agcgcgtatatagtttcctacctgtgatgatgtaggtggggctgtctcttgtccgggcaa
A H I S K D G H Y Y I H P D R E Q A R S -
ctctgaccgtaaggaagctgcaaggatccagtccttcccggataactacatatttgaag
1141 ---------+---------+---------+---------+---------+---------+ 1200
gagactggcattcccttcgacgttcctaggtcaggaagggcctattgatgtataaacttc
L T V R E A A R I Q S F P D N Y I F E G -
gatccatgacctcaaagtaccggcagataggaaacgccgtgcctccactaatgtctgaga
1201 ---------+---------+---------+---------+---------+---------+ 1260
ctaggtactggagtttcatggccgtctatcctttgcggcacggaggtgattacagactct
S M T S K Y R Q I G N A V P P L M S E K -
Ncochrom
agattgcccgcaaactttatgaaatctacagggggggacctctgatttctgaaaaggatct
1261 ---------+---------+---------+---------+---------+---------+ 1320
tctaacgggcgtttgaaatactttagatgtccccctggagactaaagacttttcctaga
Xhometilaze
Mig.MthII
L I S E K D L -
I A R K L Y E I Y R G D L *
gctgcttgatattcataccagaatctgcgaatggggaaaaaagaatttaaggaaatttcc
1321 ---------+---------+---------+---------+---------+---------+ 1380
cgacgaactataagtatggtcttagacgcttacccctttttttcttaaattcctttaaagg
L L D I H T R I C E W G K K N L R K F P -
496Nco
ctggcgcgaaagaagagatgactatggaattctgctggctgaaataatgcttcacagaac
1381 ---------+---------+---------+---------+---------+---------+ 1440
gaccgcgctttcttctctactgataccttaagacgaccgactttattacgaagtgtcttg
W R E R R D D Y G I L L A E I M L H R T -

```
     acgcgcagaacaggtgcttgaaatttatgaaaactttgttgaaaaatttcctgattttaa
1441 ---------+---------+---------+---------+---------+---------+ 1500
     tgcgcgtcttgtccacgaactttaaatacttttgaaacaacttttttaaaggactaaaatt

      R  A  E  Q  V  L  E  I  Y  E  N  F  V  E  K  F  P  D  F  K  -


     atcggtgtgtgaagcgggacaggaaacaatagaaaaagaaatggagtcccttggtctcag
1501 ---------+---------+---------+---------+---------+---------+ 1560
     tagccacacacttcgccctgtcctttgttatctttttctttacctcagggaaccagagtc

      S  V  C  E  A  G  Q  E  T  I  E  K  E  M  E  S  L  G  L  R  -


     atggagagccagaaatctccataaactcgcatgtgagattgaaagcagacacgggggggc
1561 ---------+---------+---------+---------+---------+---------+ 1620
     tacctctcggtctttagaggtatttgagcgtacactctaactttcgtctgtgccccccg

      W  R  A  R  N  L  H  K  L  A  C  E  I  E  S  R  H  G  G  A  -


     tgttccaaaaaacaaaaatgatctccttgaacttccaggcattgggaactacatttcctc
1621 ---------+---------+---------+---------+---------+---------+ 1680
     acaaggtttttttgtttttactagaggaacttgaaggtccgtaacccttgatgtaaaggag

      V  P  K  N  K  N  D  L  L  E  L  P  G  I  G  N  Y  I  S  S  -


     ggcgtttttatgtttttccaaaaatattccagaacccttctggacacgaacaccgtcag
1681 ---------+---------+---------+---------+---------+---------+ 1740
     ccgcaaaaatacaaaaaggtttttataaggtcttggggaagacctgtgcttgtggcagtc

      A  F  L  C  F  S  K  N  I  P  E  P  L  L  D  T  N  T  V  R  -


     aatcattggtagattgtttgatcttgagatcagtgattcatcccgcagaaaaaaggattt
1741 ---------+---------+---------+---------+---------+---------+ 1800
     ttagtaaccatctaacaaactagaactctagtcactaagtagggcgtctttttcctaaa

      I  I  G  R  L  F  D  L  E  I  S  D  S  S  R  R  K  K  D  F  -


     tgaaacagtaatgaggaaaattcttgaatttggtgactgcagacacctttcactatcgat
1801 ---------+---------+---------+---------+---------+---------+ 1860
     actttgtcattactccttttaagaacttaaaccactgacgtctgtggaaagtgatagcta

      E  T  V  M  R  K  I  L  E  F  G  D  C  R  H  L  S  L  S  M  -


     gatcgatttcggagaggctgtgtgcagagctagcgatcccctctgccatgaatgtcccct
1861 ---------+---------+---------+---------+---------+---------+ 1920
     ctagctaaagcctctccgacacacgtctcgatcgctaggggagacggtacttacagggga

      I  D  F  G  E  A  V  C  R  A  S  D  P  L  C  H  E  C  P  L  -


     gaaactatcatgcaacttttacaggagatgttagaatgagaaaaacaatgattgaagagg
1921 ---------+---------+---------+---------+---------+---------+ 1980
     ctttgatagtacgttgaaaat gtcctctacaatcttactctttttgttac taacttctcc

      K  L  S  C  N  F  Y  R  R  C  *     496Xho
```

7.4.2 Nucleotide sequences of genes encoding TTUDGA DNA glycosylase and neighboring X216 protein of *Thermus thermophilus* HB27

Raw sequencing data of the respective contigs from the genome sequencing project of *Thermus thermophilus* were manually edited in Staden package. The resulting consensus sequence was used for the conceptual translation into the protein sequence and determination of start and stop codons.

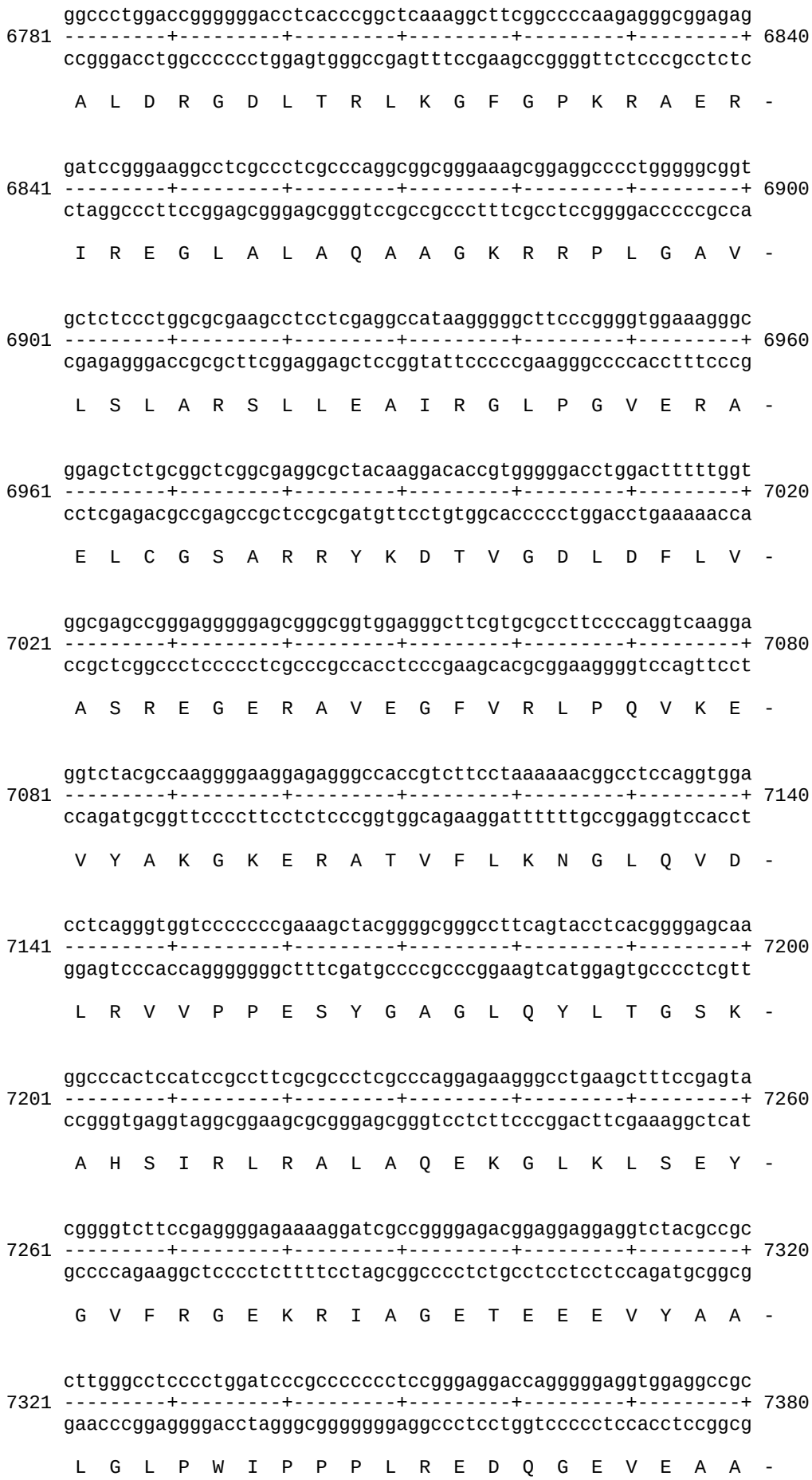



tgcttcaggcccaggcgcaaaactgcaccgcctgccgcctcgcggagggccggacccggg
3721 ---------+---------+---------+---------+---------+---------+---------+ 3780
acgaagtccgggtccgcgttttgacgtggcggacggcggagcgcctcccggcctgggccc
L Q A Q A Q N C T A C R L A E G R T R V -
tggtcttcggggagggaaacccggacgcaaagctcatgatcgtgggggaaggccccgggg
3781 ---------+---------+---------+---------+---------+---------+---------+ 3840
accagaagcccctccctttgggcctgcgtttcgagtactagcacccccttccggggcccc
V F G E G N P D A K L M I V G E G P G E –
aggaggaggacaagacgggccgccccttcgtgggcaaggcggggcagcttctaaaccgca
3841 ---------+---------+---------+---------+---------+---------+---------+ 3900
tcctcctcctgttctgcccggcggggaagcacccgttccgccccgtcgaagatttggcgt
E E D K T G R P F V G K A G Q L L N R I -
tcctggaggcggcggggatccccagggaggaggtctacatcaccaacatcgtcaagtgcc
3901 ---------+---------+---------+---------+---------+---------+---------+ 3960
aggacctccgccgcccctaggggtccctcctccagatgtagtggttgtagcagttcacgg
L E A A G I P R E E V Y I T N I V K C R -
gccccccgcaaaaccgcgcccccctttcccgacgaggccaagatctgcacggacaagtggc
3961 ---------+---------+---------+---------+---------+---------+---------+ 4020
cgggggggcgttttggcgcgggggggaagggctgctccggttctagacgtgcctgttcaccg
P P Q N R A P L P D E A K I C T D K W L -
tcctcaagcagattgagctcatcgcccccagatcatcgtccccttgggggcggtggccg
4021 ---------+---------+---------+---------+---------+---------+---------+ 4080
aggagttcgtctaactcgagtagcgggggggtctagtagcaggggaaccccgccaccggc
L K Q I E L I A P Q I I V P L G A V A A -
ccgagttcttcctgggggagaaggtctccatcaccaaggtccgggggaagtggtacgagt
4081 ---------+---------+---------+---------+---------+---------+---------+ 4140
ggctcaagaaggacccccctcttccagaggtagtggttccaggcccccttcaccatgctca
E F F L G E K V S I T K V R G K W Y E W -
ggcacgggatcaaggtcttccccatgttccaccccgcctacctcctcaggaacccgagcc
4141 ---------+---------+---------+---------+---------+---------+---------+ 4200
ccgtgccctagttccagaaggggtacaaggtggggcggatggaggagtccttgggctcgg
H G I K V F P M F H P A Y L L R N P S R -
gggccccgggaagccccaagcacctcacctggctggacatccaagaggtcaagcgggccc
4201 ---------+---------+---------+---------+---------+---------+---------+ 4260
cccggggcccttcggggttcgtggagtggaccgacctgtaggttctccagttcgcccggg
A P G S P K H L T W L D I Q E V K R A L -
tggacgccctccctcccaaggagcggcgcccggtgaaggcagtgagccaggagcccctct
4261 ---------+---------+---------+---------+---------+---------+---------+ 4320
acctgcgggagggagggttcctcgccgcgggccacttccgtcactcggtcctcggggaga
TTUDGAXho
D A L P P K E R R P V K A V S Q E P L F -

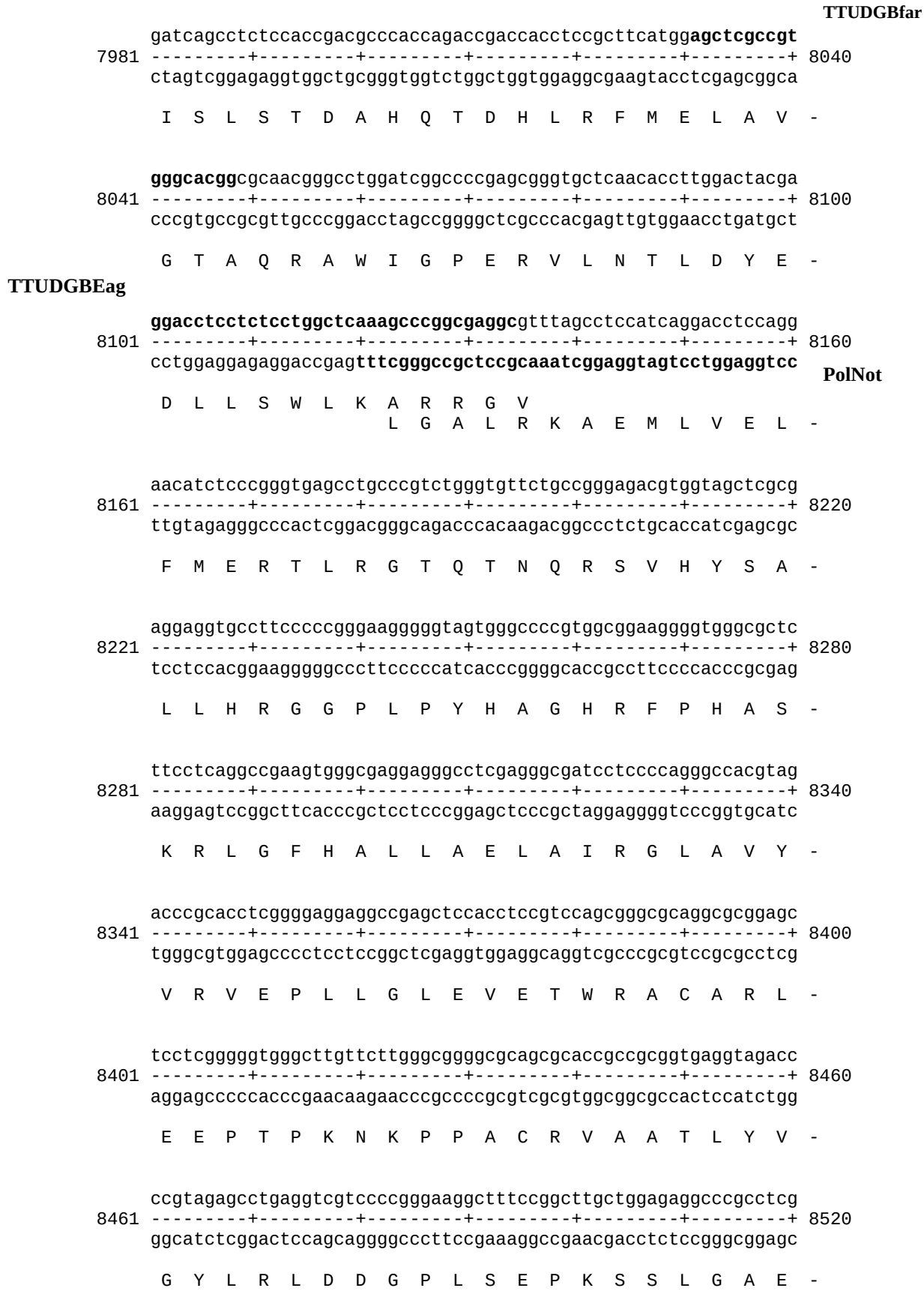
```
     tctaagcccccgcgcccaagcctcccgcccacgggaggggccctggtaggatggcctcgc
4321 ---------+---------+---------+---------+---------+---------+ 4380
     agattcgggggcgcgggttcggagggcgggtgccctccccgggaccatcctaccggagcg
```

7.4.3 Nucleotide sequence of genes encodingTTUDGB glycosylase and DNA polymerase X of *Thermus thermophilus* HB 27

```
                                                        PolNhe

     cggtccggcctcggcataggggcttcgtccagtctaccc**gtatagtgggggtatgcggaa**
6361 ---------+---------+---------+---------+---------+---------+ 6420
     gccaggccggagccgtatccccgaagcaggtcagatgggcatatcacccccatacgcctt

                                                M  R  N  -
                          DNA polymerase X  ⟶

     **ccaggagcttg**cccggatctttgaggagatcgggctcatgagcgagttcttgggggacaa
6421 ---------+---------+---------+---------+---------+---------+ 6480
     ggtcctcgaacgggcctagaaactcctctagcccgagtactcgctcaagaacccccctgtt

      Q  E  L  A  R  I  F  E  E  I  G  L  M  S  E  F  L  G  D  N  -

    ccccttccgggtccgggcctaccaccaggcggcccgaaccctctacgacctggacacccc
6481 ---------+---------+---------+---------+---------+---------+ 6540
     ggggaaggcccaggcccggatggtggtccgccgggcttgggagatgctggacctgtgggg

      P  F  R  V  R  A  Y  H  Q  A  A  R  T  L  Y  D  L  D  T  P  -

     catagaggagatcgccgaaaagggcaaggaggccctcatggagcttcccggggtggggcc
6541 ---------+---------+---------+---------+---------+---------+ 6600
     gtatctcctctagcggcttttcccgttcctccgggagtacctcgaagggccccaccccgg

      I  E  E  I  A  E  K  G  K  E  A  L  M  E  L  P  G  V  G  P  -

     ggacctcgcggagaagatcctggagttcctccgcacggggaaggtgaggaagcacgagga
6601 ---------+---------+---------+---------+---------+---------+ 6660
     cctggagcgcctcttctaggacctcaaggaggcgtgccccttccactccttcgtgctcct

      D  L  A  E  K  I  L  E  F  L  R  T  G  K  V  R  K  H  E  E  -

     actctcccagaaggtcccgcggggcgtcctcgaggtgatggaggtccccggcgtgggccc
6661 ---------+---------+---------+---------+---------+---------+ 6720
     tgagagggtcttccagggcgccccgcaggagctccactacctccaggggccgcacccggg

      L  S  Q  K  V  P  R  G  V  L  E  V  M  E  V  P  G  V  G  P  -

     caagaccgcccgtctcctctacgagggtctgggcatagactccctggagaagcttaaagc
6721 ---------+---------+---------+---------+---------+---------+ 6780
     gttctggcgggcagaggagatgctcccagacccgtatctgagggacctcttcgaatttcg

      K  T  A  R  L  L  Y  E  G  L  G  I  D  S  L  E  K  L  K  A  -
```


ggccctggaccgggggacctcacccggctcaaaggcttcggccccaagagggcggagag
6781 ---------+---------+---------+---------+---------+---------+ 6840
ccgggacctggcccccctggagtgggccgagtttccgaagccggggttctcccgcctctc
A L D R G D L T R L K G F G P K R A E R -
gatccgggaaggcctcgccctcgcccaggcggcgggaaagcggaggcccctgggggcggt
6841 ---------+---------+---------+---------+---------+---------+ 6900
ctaggcccttccggagcgggagcgggtccgccgccctttcgcctccggggacccccgcca
I R E G L A L A Q A A G K R R P L G A V -
gctctccctggcgcgaagcctcctcgaggccataaggggcttcccggggtggaaagggc
6901 ---------+---------+---------+---------+---------+---------+ 6960
cgagagggaccgcgcttcggaggagctccggtattcccccgaagggccccacctttcccg
L S L A R S L L E A I R G L P G V E R A -
ggagctctgcggctcggcgaggcgctacaaggacaccgtgggggacctggacttttggt
6961 ---------+---------+---------+---------+---------+---------+ 7020
cctcgagacgccgagccgctccgcgatgttcctgtggcacccccctggacctgaaaaacca
E L C G S A R R Y K D T V G D L D F L V -
ggcgagccgggaggggagcgggcggtggagggcttcgtgcgccttccccaggtcaagga
7021 ---------+---------+---------+---------+---------+---------+ 7080
ccgctcggccctccccctcgcccgccacctcccgaagcacgcggaaggggtccagttcct
A S R E G E R A V E G F V R L P Q V K E -
ggtctacgccaaggggaaggagagggccaccgtcttcctaaaaaacggcctccaggtgga
7081 ---------+---------+---------+---------+---------+---------+ 7140
ccagatgcggttccccttcctctcccggtggcagaaggatttttgccggaggtccacct
V Y A K G K E R A T V F L K N G L Q V D -
cctcagggtggtccccccgaaagctacggggcgggccttcagtacctcacggggagcaa
7141 ---------+---------+---------+---------+---------+---------+ 7200
ggagtcccaccaggggggctttcgatgccccgcccggaagtcatggagtgcccctcgtt
L R V V P P E S Y G A G L Q Y L T G S K -
ggcccactccatccgccttcgcgccctcgcccaggagaagggcctgaagctttccgagta
7201 ---------+---------+---------+---------+---------+---------+ 7260
ccgggtgaggtaggcggaagcgcgggagcgggtcctcttcccggacttcgaaaggctcat
A H S I R L R A L A Q E K G L K L S E Y -
cggggtcttccgaggggagaaaaggatcgccggggagacggaggaggaggtctacgccgc
7261 ---------+---------+---------+---------+---------+---------+ 7320
gccccagaaggctcccctcttttcctagcggcccctctgcctcctcctccagatgcggcg
G V F R G E K R I A G E T E E E V Y A A -
cttgggcctcccctggatcccgcccccctccgggaggaccaggggaggtggaggccgc
7321 ---------+---------+---------+---------+---------+---------+ 7380
gaacccggaggggacctagggcgggggggaggccctcctggtccccctccacctccggcg
L G L P W I P P P L R E D Q G E V E A A -

```
     cctggagggcaggcttcccaagctcctggagcttcctcaggtcaaggggacctccaggt
7381 ---------+---------+---------+---------+---------+---------+ 7440
     ggacctcccgtccgaagggttcgaggacctcgaaggagtccagttccccctggaggtcca

      L  E  G  R  L  P  K  L  L  E  L  P  Q  V  K  G  D  L  Q  V  -


     ccactccacctactccgacggccagaacaccttggaggagctctgggaagcggccaagac
7441 ---------+---------+---------+---------+---------+---------+ 7500
     ggtgaggtggatgaggctgccggtcttgtggaacctcctcgagacccttcgccggttctg

      H  S  T  Y  S  D  G  Q  N  T  L  E  E  L  W  E  A  A  K  T  -


     catgggctaccgctacctcgccgtgaccgaccactccccggcggtgcgggtggcgggggg
7501 ---------+---------+---------+---------+---------+---------+ 7560
     gtacccgatggcgatggagcggcactggctggtgaggggccgccacgcccaccgccccc

      M  G  Y  R  Y  L  A  V  T  D  H  S  P  A  V  R  V  A  G  G  -


     gccttcccccgaggaggccttgaagcgcgtggaggagatccgccgcttcaacgagaccca
7561 ---------+---------+---------+---------+---------+---------+ 7620
     cggaaggggggctcctccggaacttcgcgcacctcctctaggcggcgaagttgctctgggt

      P  S  P  E  E  A  L  K  R  V  E  E  I  R  R  F  N  E  T  H  -


     tggcccccccctacctcctcgccggggccgaggtggacatccaccccgacgggatcctgga
7621 ---------+---------+---------+---------+---------+---------+ 7680
     accgggggggatggaggagcggccccggctccacctgtaggtggggctgccctaggacct

      G  P  P  Y  L  L  A  G  A  E  V  D  I  H  P  D  G  I  L  D  -


     ctacccggactgggtcttaagggagctggacctggttttggtctccgtccactcccgctt
7681 ---------+---------+---------+---------+---------+---------+ 7740
     gatgggcctgacccagaattccctcgacctggaccaaaaccagaggcaggtgagggcgaa

      Y  P  D  W  V  L  R  E  L  D  L  V  L  V  S  V  H  S  R  F  -


     caaccttcccaaggccgaccagaccaagcgcctcctcaaggccctggaaaaccccttcgt
7741 ---------+---------+---------+---------+---------+---------+ 7800
     gttggaagggttccggctggtctggttcgcggaggagttccgggacctttggggaagca

      N  L  P  K  A  D  Q  T  K  R  L  L  K  A  L  E  N  P  F  V  -


     ccacgtcctcgcccaccccacggcgaggcttttgggccgccgcgcccccattgaggccga
7801 ---------+---------+---------+---------+---------+---------+ 7860
     ggtgcaggagcgggtggggtgccgctccgaaaacccggcggcgcgggggtaactccggct

      H  V  L  A  H  P  T  A  R  L  L  G  R  R  A  P  I  E  A  D  -


     ctgggaggccgtcttccagaaggccaaggaaaagggcgtggcggtggagattgacggcta
7861 ---------+---------+---------+---------+---------+---------+ 7920
     gaccctccggcagaaggtcttccggttccttttcccgcaccgccacctctaactgccgat

      W  E  A  V  F  Q  K  A  K  E  K  G  V  A  V  E  I  D  G  Y  -


     ctacgaccgcatggacctccccgacgacctggcccgcatggcctacgggatggggctttg
7921 ---------+---------+---------+---------+---------+---------+ 7980
     gatgctggcgtacctggaggggctgctggaccgggcgtaccggatgccctaccccgaaac

      Y  D  R  M  D  L  P  D  D  L  A  R  M  A  Y  G  M  G  L  W  -
```


TTUDGBfar
gatcagcctctccaccgacgcccaccagaccgaccacctccgcttcatggagctcgccgt
7981 ---------+---------+---------+---------+---------+---------+ 8040
ctagtcggagaggtggctgcgggtggtctggctggtggaggcgaagtacctcgagcggca
I S L S T D A H Q T D H L R F M E L A V -
gggcacggcgcaacgggcctggatcggccccgagcgggtgctcaacaccttggactacga
8041 ---------+---------+---------+---------+---------+---------+ 8100
cccgtgccgcgttgcccggacctagccggggctcgcccacgagttgtggaacctgatgct
G T A Q R A W I G P E R V L N T L D Y E -
TTUDGBEag
ggacctcctctcctggctcaaagcccggcgaggcgtttagcctccatcaggacctccagg
8101 ---------+---------+---------+---------+---------+---------+ 8160
cctggaggagaggaccgagtttcgggccgctccgcaaatcggaggtagtcctggaggtcc
PolNot
D L L S W L K A R R G V
L G A L R K A E M L V E L -
aacatctcccgggtgagcctgcccgtctgggtgttctgccgggagacgtggtagctcgcg
8161 ---------+---------+---------+---------+---------+---------+ 8220
ttgtagagggcccactcggacgggcagacccacaagacggccctctgcaccatcgagcgc
F M E R T L R G T Q T N Q R S V H Y S A -
aggaggtgccttcccccgggaaggggtagtgggccccgtggcggaaggggtgggcgctc
8221 ---------+---------+---------+---------+---------+---------+ 8280
tcctccacggaaggggggcccttcccccatcacccggggcaccgccttccccacccgcgag
L L H R G G P L P Y H A G H R F P H A S -
ttcctcaggccgaagtgggcgaggagggcctcgagggcgatcctccccagggccacgtag
8281 ---------+---------+---------+---------+---------+---------+ 8340
aaggagtccggcttcacccgctcctcccggagctcccgctaggaggggtcccggtgcatc
K R L G F H A L L A E L A I R G L A V Y -
acccgcacctcggggaggaggccgagctccacctccgtccagcgggcgcaggcgcggagc
8341 ---------+---------+---------+---------+---------+---------+ 8400
tgggcgtggagcccctcctccggctcgaggtggaggcaggtcgcccgcgtccgcgcctcg
V R V E P L L G L E V E T W R A C A R L -
tcctcgggggtgggcttgttcttgggcggggcgcagcgcaccgccgcggtgaggtagacc
8401 ---------+---------+---------+---------+---------+---------+ 8460
aggagccccacccgaacaagaacccgccccgcgtcgcgtggcggcgccactccatctgg
E E P T P K N K P P A C R V A A T L Y V -
ccgtagagcctgaggtcgtccccgggaaggctttccggcttgctggagaggcccgcctcg
8461 ---------+---------+---------+---------+---------+---------+ 8520
ggcatctcggactccagcaggggcccttccgaaaggccgaacgacctctccgggcggagc
G Y L R L D D G P L S E P K S S L G A E -

```
     tggagcaggggtagaggaaggccccggaggcgtccccggtgaaggggcggcccgtgcgg
8521 ---------+---------+---------+---------+---------+---------+ 8580
     acctcgtcccccatctccttccggggcctccgcaggggccacttccccgccgggcacgcc

      H  L  L  P  Y  L  F  A  G  S  A  D  G  T  F  P  R  G  T  R  -


     ttggagccgtgggccccgggggcgaggccgaagaggaggatcctggcctcggggtcgcca
8581 ---------+---------+---------+---------+---------+---------+ 8640
     aacctcggcacccggggcccccgctccggcttctcctcctaggaccggagccccagcggt

      N  S  G  H  A  G  P  A  L  G  F  L  L  I  R  A  E  P  D  G  -


     aagcccggaaccggcctcgcccagtagggctcgccccggaaggcccgctttctccccacg
8641 ---------+---------+---------+---------+---------+---------+ 8700
     ttcgggccttggccggagcgggtcatcccgagcggggccttccgggcgaaagaggggtgc

      F  G  P  V  P  R  A  W  Y  P  E  G  R  F  A  R  K  R  G  V  -


     acctcctcccgccaggccacgagcctggggcagaggcggcaggcggtcagggtttggacg
8701 ---------+---------+---------+---------+---------+---------+ 8760
     tggaggagggcggtccggtgctcggaccccgtctccgccgtccgccagtcccaaacctgc

      V  E  E  R  W  A  V  L  R  P  C  L  R  C  A  T  L  T  Q  V  -


     aaggcttccctgtccatcacaggggatgttgtcgtgcttcttcctgggcctctcctcct
8761 ---------+---------+---------+---------+---------+---------+ 8820
     ttcc[gaagggacaggtagtgtccccctacaacagc]acgaagaaggacccggagaggagga

      F  A  E  R  D  M
```

TTUDGBNco ← **TTUDGB**

7.4.4 Nucleotide sequence of gene encoding MutY from *Thermus thermophilus* HB 27

```
    GTGGAACCAGGCCCTCATGGAGCTCGGGGCCACGGTCTGCCTGCCGAAAAGGCCCCGCTG
601 ---------+---------+---------+---------+---------+---------+ 660
    CACCTTGGTCCGGGAGTACCTCGAGCCCCGGTGCCAGACGGACGGCTTTTCCGGGGCGAC

     W  N  Q  A  L  M  E  L  G  A  T  V  C  L  P  K  R  P  R  C  -


    CGGGACCTGTCCCCTAGGGGCCTTCTGCCGGGGGAAGGAGGCCCCCGGGCGCTACCCCGC
661 ---------+---------+---------+---------+---------+---------+ 720
    GCCCTGGACAGGGGATCCCCGGAAGACGGCCCCCTTCCTCCGGGGGCCCGCGATGGGGCG

     G  T  C  P  L  G  A  F  C  R  G  K  E  A  P  G  R  Y  P  A  -


    GCCCAGGAAGCGCTGGGCGAAGGAGGAGCGCCTCGTCGCCCTCGTCCTCCTCGGGCGGAA
721 ---------+---------+---------+---------+---------+---------+ 780
    CGGGTCCTTCGCGACCCGCTTCCTCCTCGCGGAGCAGCGGGAGCAGGAGGAGCCCGCCTT

     P  R  K  R  W  A  K  E  E  R  L  V  A  L  V  L  L  G  R  K  -


    GGGGGTGCACCTGGAAAGGCTCGAGGGCCGCTTCCAGGGCCTCTACGGCGTCCCCCTCTT
781 ---------+---------+---------+---------+---------+---------+ 840
    CCCCCACGTGGACCTTTCCGAGCTCCCGGCGAAGGTCCCGGAGATGCCGCAGGGGGAGAA

     G  V  H  L  E  R  L  E  G  R  F  Q  G  L  Y  G  V  P  L  F  -


    TCCCCCTGAGGAGCTTCCCGGGCGGGAGGCGGCCTTCGGGGTGAGGTCTAGGCCCCTAGG
841 ---------+---------+---------+---------+---------+---------+ 900
    AGGGGGACTCCTCGAAGGGCCCGCCCTCCGCCGGAAGCCCCACTCCAGATCCGGGGATCC

     P  P  E  E  L  P  G  R  E  A  A  F  G  V  R  S  R  P  L  G  -


    CGAGGTGCGCCACGCCCTCACCCACCGGAGGCTTCTCGTGGAGGTGCGGGGGGCCCTTTG
901 ---------+---------+---------+---------+---------+---------+ 960
    GCTCCACGCGGTGCGGGAGTGGGTGGCCTCCGAAGAGCACCTCCACGCCCCCCGGGAAAC

     E  V  R  H  A  L  T  H  R  R  L  L  V  E  V  R  G  A  L  W  -


    GGAAGGGGAGGGGGAGGACCCCTGGAGGAGGCCCCTACCCAAGCTCATGGAGAAGGTGCT
961 ---------+---------+---------+---------+---------+---------+ 1020
    CCTTCCCCTCCCCCTCCTGGGGACCTCCTCCGGGGATGGGTTCGAGTACCTCTTCCACGA

     E  G  E  G  E  D  P  W  R  R  P  L  P  K  L  M  E  K  V  L  -


    CCGCAAGGCGCTGCCCCTCCTCGCTCATGCGGGCGTAGTCCCCCTCCCGGACGCATAGGC
1021 ---------+---------+---------+---------+---------+---------+ 1080
    GGCGTTCCGCGACGGGGAGGAGCGAGTACGCCCGCATCAGGGGGAGGGCCTGCGTATCCG

     R  K  A  L  P  L  L  A  H  A  G  V  V  P  L  P  D  A        thermusHind


    CCACGGCGTATAGCCCTTCCAGGCGCTTGAGGCGGAAGGTGCCTTCCTCCCACTCCTCGG
1081 ---------+---------+---------+---------+---------+---------+ 1140
    GGTGCCGCATATCGGGAAGGTCCGCGAACTCCGCCTTCCACGGAAGGAGGGTGAGGAGCC
```

7.5 Acknowledgements

My particular gratitude goes to my supervisor, Prof. Dr. Hans-Joachim Fritz for accepting me as a PhD student in his research group, involving into the fascinating field of DNA repair and numerous stimulating discussions during the work.

I am thankful to Dr. Wilfried Kramer for being co-referent of the thesis and for the valuable remarks on the theoretical and practical aspects of the work.

My appreciation to Dr. Harald Kolmar for well-aimed and pictureous solutions to various problems. Many thanks to Dr. Rainer Merkl for taking care of some bioinformatic aspects of this work.

My gratefulness to Dr. Yvonne Fondufe - Mittendorf for being an excellent colleague and adviser, also for the courage and patience in reading and correcting this manuscript. My warm acknowledgements to Dr. Susanne Behrens for the friendly shearing of the working space and numerous advices.

Many thanks to past and present members of the Genome Analysis Laboratory, particularly to Dr. Hans-Peter Klenk, Dr. Carsten Jacobi, Dr. Thomas Hartsch and Dr. Anke Henne in helping to decipher the secrets of the genome of *Thermus thermophilus* and the support with laboratory material.

Thanks goes to Dr. Uwe Deppenmeier and Dr. Hans-Jörg Abken from Institute Microbiology and Genetics for the help in cultivating *Methanobacterium thermoautotrophicum*, for some laboratory material and relevant literature.

Silke Schlenczeck deserves particular attention and warm words for the invaluable technical help and patience, which significantly enriched this work.

My appreciation to Olaf Waase for fixing of various possible and sometimes impossible technical problems. Many thanks to Marita Kalck and Kirsten Hahne for guiding through bureaucratic labyrinths.

Special thanks to Andreas Christman for the help with computer work and, particularly, in poster making and printing. To Dr. Rüdiger Dietrich for help in western - blots and for fruitful philosophical discussions, to Martina Henn-Sax for brainstorming sessions about the enzyme kinetics, to Christian Rudolph for fixing always appearing computer problems, to Anke Schürer for solving a lot of DNA repair and basically all other problems, to Dr. Martin Laging for instructive lesions about BioCad station and numerous purified DNA oligonucleotides, to Julia Wittmann for the help in visualizing mechanisms of glycosylase action. Many thanks to Gülden Aral for jet speed in supplying with urgent materials, to Marlies Bartel and Angelika Löffel for readiness to help in organising matters of everyday work.

I am thankful to all my colleagues, Rene, Cristian, Eva, Christian, Milena, Stina, Dennis, Alexander, Frank, Corina, Haiquan, Cristina, Martina for friendliness, support and refreshing conversations around the cup of coffee.

Finally, my profound gratitude goes to my family and friends for their support and understanding.

7.6 Curriculum vitae

Personal details:

Name: Vytaute Starkuviene

Date and place of birth: 6th March 1970 in Marijampole, Lithuania

Nationality: lithuanian

Educational background:

1977-1988 Primary and secondary school No 50 of Vilnius, Vilnius, Lithuania

1988-1992 Studies of biochemistry, University of Vilnius, Lithuania

1992-1993 Diploma work at Biotechnology Institute "Fermentas" under the supervision of Dr. S. Zvirblis. The title: "Purification and characterization of D-hydantoinase from *Bacillus circulans*".

Chronology of employment

1993-1995 Research assistant in "Protein Structure and Interaction Laboratory", Biotechnology Institute "Fermentas", Vilnius, Lithuania.

1995-1997 Research assistant in Human Genetics Center, Vilnius, Lithuania.

1997-1998 Visiting Scientist in "Biochemical Instrumentation Program", in the laboratory of Prof. W. Ansorge, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.

June 1998 till date: Ph.D studies at the Institute of Microbiology and Genetics, Georg-August University Göttingen under the supervision of Prof. Dr. H-J. Fritz