Chapter 1

General introduction into Nudix hydrolase proteins

Different chemical and physical forces constantly threaten the integrity of the genetic code in living organisms. These organisms are therefore equipped with special error avoidance pathways enabling them to retain their original genetic status. Oxidative DNA damage results from exposure to both endogenous and exogenous oxidizing agents, e.g. reactive oxygen molecules such as superoxide, hydrogen peroxide and hydroxyl radicals. These compounds are formed in vivo as byproducts of normal aerobic metabolism. Although cells possess many defense mechanisms against oxidative damage, it has been estimated that the genome of a mammalian cell receives about 10^4 - 10^5 oxidative hits per day (Kreutzer and Essigmann, 1998).

Base substitutions are by far the most frequent occurring mutation (70.8%), followed by deletions (17.2%), duplications (7.7%), and single base frame shifts (4.3%). The most frequently occurring base substitution mutation observed in aerobic organisms is a GC to AT transition. Oxidized, deaminated cytosines are thought to be the major source of GC to AT transitions, because an oxidized uracil species can be formed from a cytosine base. This implies that through deamination of the 4-amino group of cytosine, a cytosine-like base-pairing moiety can be replaced by a thymine-like base-pairing moiety (Kreutzer and Essigmann, 1998).

Base substitutions as a result of transversions occur with a lower frequency than transitions. The base transversion observed most frequently is that of AT to CG. Treffers (Treffers *et al.*, 1999) originally described a mutant strain of *E. coli*, *mutT*1, with mutation frequencies ranging from 100- to 10,000-fold higher than normal. The mutations occurring in this strain were later shown to be specifically increasing the rate of AT to CG transversions (Yanofsky *et al.*, 1966). The *mutT* gene was found to encode a 129 amino acid protein with a Mr of 15 kDa (Xia *et al.*, 1992), capable of hydrolyzing an oxidized form of dGTP, 7,8-dihydro-8-oxo-deoxyguanine (8-oxo-dGTP). When used as a substrate for DNA synthesis, this nucleotide is efficiently inserted opposite of a dAMP residue, thus causing AT to CG transversions (Maki and Sekiguchi, 1992) (See below).

The MutT or Nudix hydrolase protein family

In an amino acid comparison of the *E. coli* MutT sequence with sequences in databases using the BLAST program (Altschul *et al.*, 1990), Koonin *et a*l. discovered limited similarities with putative proteins of uncharacterized ORFs of viruses, pro- and eukaryotic organisms. All these similarities consistently highlight the same amino acid residues of the MutT protein ($GX_5EX_7REX_2EEXG$), suggesting their possible functional importance (Fig. 1). (Koonin, 1993) Identification and characterization of enzymatic activities of different members of this protein family has shown that this motif is not only found in MutT proteins of different organisms (Akiyama *et al.*, 1987;Mejean *et al.*, 1994;Kamath and Yanofsky, 1993;Sakumi *et al.*, 1993;Kakuma *et al.*, 1995;Cai *et al.*, 1995), but also in proteins with functions quite distinct from MutT (O'Handley *et al.*, 1996;Frick and Bessman, 1995;Hurtado *et al.*, 1987;O'Handley *et al.*, 1998;Frick *et al.*, 1995a;Sheikh *et al.*, 1998;Conyers and Bessman, 1999;Raffaelli *et al.*, 1999;Frick *et al.*, 1995a;Safrany *et al.*, 1998;Espinosa *et al.*, 1999;Safrany *et al.*, 1999). However, they share one common feature, namely the hydrolysis of a pyrophosphate bond of a **Nu**cleoside **di**phosphate compound linked to some other moiety, **X**. These proteins are therefore also being referred to as **Nudix** hydrolases (Bessman *et al.*, 1996).

Analysis of the different genome sequences indicates that Nudix hydrolases are not only ubiquitously spread in nature, but also that multiple members can be found in a single organism, varying in predicted protein sizes. *E. coli* for instance contains 11 open reading frames (ORFs) (putatively) encoding Nudix hydrolase proteins (Sheikh *et al.*, 1998). A selected overview of the occurrence of Nudix hydrolase proteins encoded by the genomes of various organisms and the size of those (putative) proteins is presented in Table 1.

Organism	Accession number	Partial sequence	Gene	Substrate of protein
E. coli	P08337	EFPG**G**KIEMG**E**TPEQAVV**RE**LQ**EE**V**G**ITPQHFSLFEKLEYEFPDRH	*mutT*	8-oxo-dGTP
H. influenza	I64101	EFPG**G**KVDAG**E**TPEQALK**RE**LE**EE**I**G**IVALNAELYERFQFEYPTKI	*mutT*	8-oxo-dGTP
P. vulgaris	P32090	EFPG**G**KLEDN**E**TPEQALL**RE**LQ**EE**I**G**IDVTQCTLLDTVAHDFPDRH	*mutT*	8-oxo-dGTP
Str.pneumoniae	P41354	IGVG**G**KLERG**E**TPQECAV**RE**IL**EE**T**G**LKAKPVLKGV---ITFPEFT	*mutX*	8-oxo-dGTP
Human	P36639	NGFG**G**KVQEG**E**TIEDGAR**RE**LQ**EE**S**G**LTVDALHKVG--QIVFEFVG	*Hmth1*	8-oxo-dGTP
E. coli	P36651	EMVA**G**MIEEG**E**SVEDVAR**RE**AI**EE**A**G**LIVKRTKPVLS-FLASPGGT	*Yqie*	ADPR
B. subtilis	P54570	EIPA**G**KLEKG**E**EPEYTAL**RE**LE**EE**T**G**YTAKKLTKITA-FYTSPGFA	*Yqkg*	ADPR
H. influenza	P44684	ELIA**G**MVEKG**E**KPEDVAL**RE**SE**EE**A**G**IQVKNLTHCLS-VWDSPGGI	*yqie*	ADPR
B.methanolicus	AY128667	EIPA**G**KLEKG**E**DPRVTAL**RE**LE**EE**T**G**YECEQMEWLIS-FATSPGFA	*Act*	ADPR
S. ambofaciens	CAA06447	ELPG**G**VLELD**E**TPETGVA**RE**VW**EE**T**G**IRVEVDELTGVYKNTTRGIV	*Orf131*	n.d.
A. methanolica	L36679	EVPG**G**IIDGD**E**SPEETVV**RE**IE**EE**T**G**YRPRSIEPLITFEPAVGMLR	*orf192*	ADPR
E. coli	P45799	GFSK**G**LIDPG**E**SVYEAAN**RE**LK**EE**V**G**FGANDLTFLKK-LSMAPSYF	*orf186*	Ap3A
Human	P50583	TPPK**G**HVEPG**E**DDLETAL**RE**TQ**EE**A**G**IEAGQLTIIEGFKRELNYVA	*nudt2*	Ap4A
E. coli	P32664	TVLA**G**FVEVG**E**TLEQAVA**RE**VM**EE**S**G**IKVKNLRYVTS--QPWPFPQ	orf257	NADH
E. coli	P24236	QSVT**G**SVEEG**E**TAPQAAM**RE**VK**EE**VTIDVVAEQLTLIDCQRTVEFE	*orf17*	dATP
Human	NP006694	IVPG**G**GMEPE**E**EPSVAAV**RE**VC**EE**A**G**VKGTLGRLVGIFENQERKHR	*dipp*	PP-InsP_5
E.coli	AAC77844	FVPG**G**RVQKD**E**TLEAAFE**R**LTMA**E**L**G**LRLPITAGQFYGVWQHFYDD	*orf1.9*	GDP-sugar
Consensus		**G E RE EE G**		

Figure 1. Partial alignment of proteins carrying the Nudix hydrolase motif (bold characters). Dipp: Diphosphoinositol polyphosphate; ADPR: ADP ribose; $Ap_{3/4}A$: diadenosine tri/tetraphosphate; PP-InsP_5: diphosphopentakiphosphate; n.d.: not determined.

Table 1. The number of Nudix hydrolase encoding genes found in the genomes of some selected organisms, the number of genes with Nudix hydrolase motif related to genome size and the range of the predicted sizes of the (putative) proteins. Results were obtained using the Blast protocol (Altschul *et al.*, 1990) on the genomes of the specified organisms.

Organism	Number of genes with Nudix hydrolase motif	Genome size (Mb)	Number of genes with Nudix hydrolase motif related to genome size (genes/Mb)	Range of predicted sizes of protein product (a.a.)
Escherichia coli K12	11	4.6	2.4	129-257
Bacillus subtilis	4	4.2	1.0	149-185
Bacillus anthracis	30	5.2	5.8	108-205
Bacillus cereus	26	5.4	4.8	124-212
Lactococcus lactis ssp. *lactis*	8	2.3	3.4	146-194
Streptomyces coelicolor	20	8.7	2.3	128-359
Streptomyces avermitilis	20	9.0	2.2	130-346
Mycobacterium tuberculosis	9	4.4	2.0	141-351
Clostridium acetobutylicum	7	3.9	1.8	146-307
Deinococcus radiodurans	21	3.3	6.3	91-548
Methanococcus jannaschii	1	1.7	0.6	169
Saccharomyces cerevisiae	4	13	0.3	189-971
Caenorhabditis elegans	5	97	0.05	150-365
Arabidopsis thaliana	7	157	0.04	147-304
Human	19	3200	0.006	140-1503

Although these enzymes are highly disparate in their substrate specificity, it has been suggested that a common feature is shared by all of these enzymes: they hydrolyze potentially hazardous compounds or they prevent the unbalanced accumulation of certain metabolites. In that sense Nudix hydrolases are described as proteins involved in modulating the accumulation of intermediates in biochemical pathways or as "housecleaning" enzymes (Bessman *et al.*, 1996). More recently, MutT motif containing proteins have been characterized which do not entirely

correspond to the characteristics of Nudix hydrolases, neither in their function (chapter 3 and 5) nor in their substrate specificity (Safrany *et al.*, 1998). The methanol dehydrogenase (MDH) activator protein (ACT) of *Bacillus methanolicus* was found to stimulate MDH activity by hydrolyzing a tightly bound NADH cofactor, an activity that doesn't match the description "housecleaning" enzyme (chapter 3). Orf192, a protein encoded by the *Amycolatopsis methanolica* indigenous plasmid pMEA300, is involved in the regulation of both plasmid replication and the plasmid-encoded pock formation phenotype (Chapter 5). Diphosphoinositol polyphosphate phosphohydrolase (DIPP) is the first example of a MutT motif-containing enzyme hydrolyzing a non-Nudix compound. Rat hepatic DIPP was purified and characterized as an enzyme cleaving the β-phosphate from the diphosphate groups in diphosphoinositol pentakiphosphate (PP-$InsP_5$) and bisdiphosphoinositol tetrakiphosphate ($[PP]_2$-$InsP_4$) (Safrany *et al.*, 1998). Whether these proteins are the exceptions that prove the rule remains to be seen.

Even though various substrates have been identified for Nudix hydrolase enzymes, the functions of only a few of those proteins have been described in more detail, linking their activities to specific metabolic or regulatory steps. Apart from the ACT protein of *B. methanolicus* MDH, two other Nudix hydrolase proteins have been functionally characterized: the human and yeast mRNA decapping enzyme and the human calcium channel LTRPC2, which is gated by the Nudix compound ADP-ribose. Eukaryotic mRNA molecules contain a 7-methyl-GTP at the 5' end and a poly-A tail at the 3' terminus. Degradation of mRNA –playing a key role in the regulation of gene expression- proceeds via removal of both the poly-A tail and decapping of the 7-methyl-GTP moiety of mRNA. The proteins responsible for the decapping activity in yeasts and humans, the Nudix hydrolase motif containing Dcp2 proteins, have been characterized, revealing their ability to remove 7-methyl-GDP from capped mRNA. Orthologs of these proteins were also identified in other eukaryotes, all of them containing the Nudix hydrolase motif (Van Dijk *et al.*, 2002;Wang *et al.*, 2002).

Another example of a well-characterized protein containing the Nudix hydrolase motif is the human LTRPC2 (or TRPM2) protein, which is an ADP-ribose dependent calcium-permeable cation channel. In cells expressing this cation channel, it was found that intracellular ADP-ribose regulates entry of calcium into these cells in response to oxidative stress (Perraud *et al.*, 2001). Experiments using LTRPC2 with a deleted Nudix motif, confirmed the ADP-ribose-induced gating of this cation channel. Furthermore, stress-induced gating could be inhibited by pharmacological reagents that inhibit NAD hydrolysis to ADP-ribose, combined with cytosolic or mitochondrial overexpression of an enzyme that specifically hydrolyzes ADP-ribose in the HEK host cells (Perraud *et al.*, 2005).

Nudix hydrolase enzymes of which only the substrate specificities have been characterized and which were generically typified as "housecleaning" enzymes thus may have important metabolic or regulatory roles which largely remain to be elucidated.

MutT proteins

MutT, the first well characterized Nudix hydrolase protein, was first shown to exhibit a weak dGTP hydrolyzing activity, according to the equation dGTP → dGMP + PPi (Bhatnagar *et al.*, 1991;Bhatnagar and Bessman, 1988). The reaction strictly required magnesium ions for its activity and a sharp pH optimum was observed at pH 9.0 (Bhatnagar *et al.*, 1991).

Maki en Sekiguchi (Maki and Sekiguchi, 1992) showed that labeled dGTP could be incorporated onto a poly (dA)/oligo$(dT)_{20}$ template by the action of *E. coli* DNA polymerase III and the misincorporated nucleotide was not removed by the editing function of the enzyme. However, misincorporation was specifically prevented in the presence of MutT protein. Thin-layer chromatography analysis of the misincorporated dGTP revealed that an aberrant form of dGTP had been incorporated, namely 7,8-dihydro-8-oxo-deoxyguanine (8-oxo-dGTP or GO).

This compound is produced during normal cellular metabolism by the action of free oxygen radicals, and is potentially mutagenic because of its ambiguous pairing with cytosine and adenine (Michaels and Miller, 1992). It was shown that 8-oxo-dGTP was efficiently inserted opposite a dA residue as well as a dC residue present on the DNA template (Fig. 2). The apparent K_m for the hydrolysis of 8-oxo-dGTP was 2,000 times lower than for dGTP, which led to the conclusion that 8-oxo-dGTP was the substrate for MutT. Hydrolysis of this mutagenic substrate thus prevents AT to CG transversions (Maki and Sekiguchi, 1992).

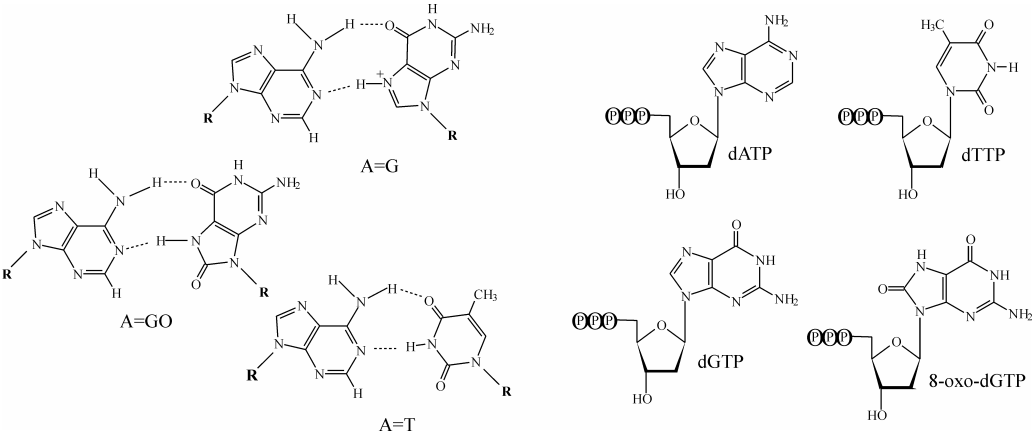

Figure 2. Base pairing and mispairing between A-T, A-G and A-GO and the structures of 8-oxo-dGTP (GO), dATP, dTTP and dGTP.

MutT protein is thus part of the so called GO-system, an error avoidance pathway in *E. coli* that is composed of at least three proteins: MutM, MutY and MutT. While MutT is concerned with the removal of the oxidatively damaged form of the guanine base from the nucleotide pool, MutM and MutY are involved in the removal and correction of GO mutations in DNA (Michaels and Miller, 1992).

Structure and catalytic mechanism of *E. coli* MutT

The hydrolysis of (8-oxo-)dGTP by MutT proceeds with a nucleophilic attack of a water molecule (or -less likely- a nucleophilic group on the enzyme) on the β-phosphorus atom of dGTP resulting in bond cleavage between the β-phosphorus atom and the α,β bridging oxygen atom, and with the departure of dGMP as the leaving group (Weber *et al.*, 1992) (Fig. 3). Only a small group of enzymes catalyze nucleophilic substitutions at the β-phosphorus atom of nucleoside triphosphates. In this group of enzymes MutT is the only hydrolase, while the other members are all synthetases (Weber *et al.*, 1992). MutT requires two divalent cations (Mg^{2+} or Mn^{2+}) for catalytic activity. There may be four possible roles of this quaternary MutT-M^{2+}-GTP-M^{2+} complex: first, a further activation of the electron rich β-phosphorus atom for the nucleophilic attack; second, to facilitate the departure of the leaving group, dGMP; third, coordination and activation of the attacking nucleophile (the water molecule); fourth, the bridging metal ion may adjust the protein conformation, bringing appropriate catalytic residues into contact with both substrates (Frick *et al.*, 1994).

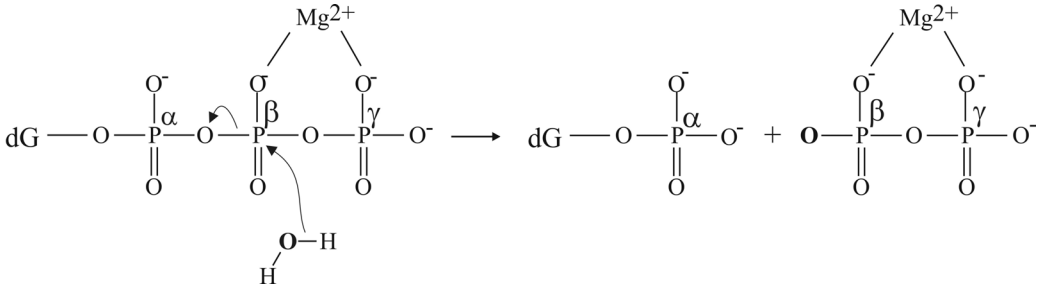

Figure 3. Proposed catalytic mechanism of the *E. coli* MutT reaction (Weber *et al.*, 1992) showing the nucleophilic attack of a water molecule on the β-phosphorus atom of dGTP.

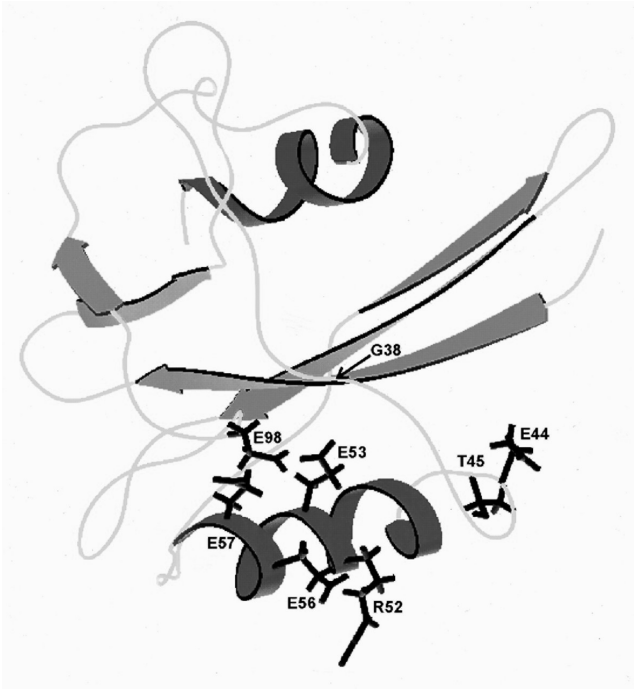

Figure 4. Ribbon diagram of *E. coli* MutT showing amino acids of the conserved motif (Bessman *et al.*, 1996)

The secondary structure of MutT determined by NMR revealed that it consists of two α-helices and a mixed β-sheet containing 5 β-strands (Fig. 4) (Weber *et al.*, 1993;Frick *et al.*, 1995b). Small proteins that contain mixed β-sheets show unusually high stability (Weber *et al.*, 1993). Amino acid residues belonging to the strongly conserved Nudix hydrolase motif are located in loop I, helix I and the start of loop II. The interactions of dGTP, adenosine 5'-α,β-methylenetriphosphate (AMPCPP) -a non-hydrolysable substrate analogue- and divalent cations with MutT, as determined by NMR, provided the first biophysical evidence of the active site being located in this region of the enzyme. At least one of the two essential divalent cations is bound in the proximity of Gly-37, Gly-38, Lys-39 and Glu-57, while the nucleotides were shown to bind near residues Leu-54 and Val-58 by NMR relaxation methods (Frick *et al.*,

1995b).

The complete tertiary structure of MutT, determined by NMR with an overall stereochemical quality corresponding to a 2.5 Å X-ray structure, shows these residues to be near each other. Furthermore, a cluster of five glutamate residues (E41, E53, E56, E57 and E98) forms a patch of strongly negative electrostatic potential likely constitutes the metal binding site (Fig. 4). A deep cleft adjacent to this site between β-strands A, C and D and loop I may form part of the nucleotide binding site (Abeygunawardana *et al.*, 1995). This location of the nucleotide binding site on the MutT protein confirms mutational analysis performed previously on the *Streptococcus pneumoniae* MutX protein (Mejean *et al.*, 1994).

A more complete elucidation of the reaction mechanism of MutT came by the solution of the structure of the quaternary MutT-Mg^{2+}- AMPCPP-Mg^{2+} complex by NMR. The fold of the enzyme in complex is quite similar to that of the free enzyme. One of the water ligands of the enzyme-bound metal ion is well positioned to attack the β-phosphorus atom. Lys-39 is positioned to interact electrostatically with the α-phosphoryl group and will facilitate the departure of the GMP leaving group (Lin *et al.*, 1997).

MTH1: the human MutT analog

Since the isolation and characterization of the *E. coli* MutT protein, a number of MutT homologs have been found in prokaryotic (Mejean *et al.*, 1994;Kamath and Yanofsky, 1993) and eukaryotic organisms (Sakumi *et al.*, 1993;Kakuma *et al.*, 1995;Cai *et al.*, 1995). All of them were capable to complement $MutT^-$ mutant cells of *E. coli*. Apart from the *E. coli* protein, the human counterpart of MutT, MTH1 (MutT homolog1), has been studied most extensively . MTH1 is a monomeric protein with a Mr of 18 kDa. Like the *E. coli* enzyme, MTH1 preferentially hydrolyzes 8-oxo-dGTP over dGTP and dATP, although the affinity of the enzyme for dGTP is 70 times lower than for 8-oxo-dGTP (Mo *et al.*, 1992). However, contrary to *E. coli* MutT, 8-oxo-dGTP is not the only mutagenic substrate for MTH1. Also 2-oxo-dATP and 8-oxo-dATP are hydrolyzed by MTH1, the first even with a two times higher catalytic efficiency ($kcat/K_m$) than 8-oxo-dGTP (Fujikawa *et al.*, 1999). Also 2-oxo-dATP is formed abundantly by reactive oxygen species in cells and elicits GC to AT transversions (Inoue *et al.*, 1998). As a result of alternative splicing of the *mth1* mRNA, two or three other MTH1 proteins have been shown to originate from the *mth1* gene (Fujikawa *et al.*, 1999;Oda *et al.*, 1997). It remains possible that other forms of the human MTH1 protein have substrate specificities different from the one characterized (Fujikawa *et al.*, 1999).

Another disparity between MTH1 and *E. coli* MutT is the higher activity of MutT for 8-oxo-GTP. While the *E. coli* enzyme catalyzes 8-oxo-rGTP and 8-oxo-dGTP cleavage with the same efficiency, the rate of hydrolysis of 8-oxo-rGTP by MTH1 is 50 times lower than its activity on 8-oxo-dGTP (Taddei *et al.*, 1997;Hayakawa *et al.*, 1999). In this regard it may be speculated that the hydrolysis of 8-oxo-rGTP in mammalian cells is carried out by another enzyme to prevent mismatches to occur in RNA (Hayakawa *et al.*, 1999). However, no such activity has been found yet. Northern blot analyses on human tissue showed that *mth1* mRNA is present in virtually all tissue types, although the rate of expression differs with a factor 20. Tissues with the most abundant *mth1* expression are testis and thymus. However, expression levels of *mth1* mRNA in Jurkat cells, a human T cell leukemia cell line, exceeded these tissues with a factor 3 (Oda *et al.*, 1997).

Targeted disruption of the MTH1 gene homolog in mice resulted in a twofold increase in tumor development during a normal life span, compared to MTH1 proficient mice. Although these entities are difficult to compare, the authors state that this mutation factor is considerably lower than the 100-fold increased mutation rate in $mutT^-$ *E. coli* mutants (Tsuzuki *et al.*, 2001). They suggest the presence of a redundant enzyme system in mice. A mouse cDNA clone was

identified that considerably suppressed the high mutation rate in an *E. coli* MutT-negative mutant strain (Tsuzuki *et al.*, 2001). Recent characterization of the encoded protein -referred to as MTH2- revealed that it contained the Nudix hydrolase motif and could hydrolyze 8-oxo-dGTP. When expressed in an *E. coli* mutT$^-$ mutant strain it was shown to reduce significantly its elevated level of spontaneous mutations (Cai *et al.*, 2003).

Non-MutT protein Nudix hydrolases in *E. coli*

Many Nudix hydrolase protein family members have now been identified that hydrolyze a substrate other than 8-oxo-dGTP. Beside *mut*T*, E. coli* contains ten other Nudix hydrolase encoding ORFs (Sheikh *et al.*, 1998). Eight of these enzymes have now been characterized, all of them showing a different substrate specificity. The first of these was *orf17* of *E. coli* which is located just upstream of the *ruvC* gene encoding a holliday-junction specific endonuclease (O'Handley *et al.*, 1996;Sharples and Lloyd, 1991). The 17 kDa protein encoded by *orf17* hydrolyzes all of the canonical (deoxy-) nucleotides, although the hydrolysis of dATP is catalyzed preferentially. The catalytic efficiency of the enzyme with this substrate is about one order of magnitude higher than with the other deoxy-nucleotides. Pyrophosphate strongly inhibits the dATP hydrolysis reaction (K_i = 20 μM). In contrast to Orf17-dATPase, MutT is hardly affected by pyrophosphate product inhibition; it has a 50-fold higher K_i, while it shares many of the properties of the Orf17 enzyme (O'Handley *et al.*, 1996). The *in vivo* role of Orf17 remains to be elucidated. No mutator phenotype could be assigned to the protein. Overproduction of Orf17 in a complementation experiment of an *E. coli mutT*$^-$ mutant strain did not decrease the mutation frequency. Deletion of *orf17* did not significantly enhance the mutation frequency when compared to its wild type parent (O'Handley *et al.*, 1996).

The protein product of *orf257* (*yjaD*), a second *E. coli* Nudix hydrolase not involved in 8-oxo-dGTP hydrolysis, catalyzes the hydrolysis of the pyrophosphate bond in a broad range of dinucleotide pyrophosphates, with a clear preference for NADH (Frick and Bessman, 1995). There doesn't seem to be any obvious use for an enzyme hydrolyzing NADH, which is an important cofactor/coenzyme in many cellular processes. Frick *et al*. (Frick and Bessman, 1995) argue that it may regulate the intracellular NADH/NAD-ratio, which is known to be an important factor in maintaining a balance between anabolic and catabolic pathways in higher organisms. Homologues of Orf257 are also found in other organisms, such as *Deinococcus radiodurans, Mycobacterium tuberculosis*, various bacteria of the gamma subdivision of proteobacteria, yeasts and humans, in which they were identified as multi domain proteins. The N-terminal part of those proteins contains a Zinc-ribbon domain, which is probably involved in DNA binding (Makarova *et al.*, 2000).

The *E. coli orf186* gene encodes a Nudix hydrolase with a rather broad substrate specificity, with its major substrates being diadenosine triphosphate (Ap_3A), ADP-ribose and NADH (O'Handley *et al.*, 1998). Diadenosine oligophosphates (Ap_nA) are intracellular and extracellular signaling molecules present in the whole spectrum of organisms from bacteria to higher eukaryotes. In prokaryotes, heat shock and oxidative stress cause accumulation of Ap_nA, which bind to and inhibit the oxidative stress-related proteins (Kisselev *et al.*, 1998). Similar to the *orf186* gene, *ygdP* of *E. coli* encodes a diadenosine oligo phosphate hydrolase, mainly specific for Ap_5A and Ap_6A. The reaction products formed from the Ap_5A substrate are ATP and ADP. YgdP is associated with the invasiveness of *E. coli* K1, causing neonatal meningitis (Bessman *et al.*, 2001). An Ap_nA hydrolyzing enzyme was also found in the human pathogen *Bartonella bacilliformis*, the only organism known that invades erythrocytes and causes human Oroya fever. *E. coli* could gain the invading capability when transformed with a plasmid containing a 1.5 Mb fragment of the *B. bacilliformis*. Analysis of this fragment revealed the presence of two ORFs, *ialA* and *ialB* (Mitchell and Minnick, 1995). The *ialA* encoded protein

belongs to the Nudix hydrolase protein family and is similar to plant diadenosine tetraphosphate hydrolases, as found in lupinus and hordeum (Maksel *et al.*, 1998). Characterization revealed that it is capable to hydrolyze Ap_4A, Ap_5A and Ap_6A, but with a clear preference for the former (Conyers and Bessman, 1999;Cartwright *et al.*, 1999). ATP always is one of the products formed in the hydrolysis reaction, irrespective of the substrate used. This indicates that the nucleophilic attack of the enzyme is directed towards the γ-phosphorus atom and not to the one on the β position, as seen with MutT of *E. coli* (Weber *et al.*, 1992;Conyers and Bessman, 1999). Homologs of this protein are also found in other invasive pathogens and is it speculated that they may function to reduce stress-induced dinucleotide levels during the invasion and so enhance pathogen survival (Cartwright *et al.*, 1999;Gaywee *et al.*, 2002).

Virtually all Nudix hydrolase proteins characterized so far were reported to hydrolyze pyrophosphate bonds in (di-) nucleotide substrates. An exception to that rule is the protein encoded by *orf1.9* of *E. coli*, which hydrolyses its substrates, GDP-mannose or GDP-glucose, to GDP and mannose or glucose. It can therefore be referred to as a GDP-mannose mannosyl hydrolase (GDPMH). Like the other Nudix hydrolases, its catalytic activity is strictly divalent ion dependent, and it has an alkaline pH optimum (pH 9.3) (Frick *et al.*, 1995a). Two strongly conserved glutamate residues within the Nudix hydrolase consensus motif REX_2EE are missing in the GDPMH sequence (Fig. 1). NMR studies indicated that, unlike all other Nudix hydrolase proteins characterized so far, GDPMH catalyzes a nucleophilic substitution at the C1' carbon of the sugar rather than at the β-phosphorus atom of the pyrophosphate moiety of the GDP-sugar (Legler *et al.*, 2000;Legler *et al.*, 2002).

Not only *E. coli* Orf186 is capable to hydrolyze ADP-ribose, also the gene product of the *E. coli orf209* was shown to use this compound as a substrate. While the substrate specificity of Orf186 was rather low, the homodimeric Orf209 protein has a clear preference for ADP-ribose. Sequence comparison of enzymes characterized as ADP-ribose hydrolases revealed the presence of a conserved proline residue 17-18 amino acids downstream of the REX_2EE motif in these enzymes, which may be useful for classification of uncharacterized Nudix hydrolase proteins. Similarly, Ap_nA hydrolases were found to contain a conserved tyrosine 18-20 amino acids downstream of this motif (Dunn *et al.*, 1999) (Fig. 1). Elucidation of the 3D structure of the *orf209* encoded ADP-ribose hydrolase in complex with Mg^{++} and a non-hydrolysable substrate analogue (α,β-methylene ADP-ribose), revealed that free enzyme is in the open conformation, while binding of Mg^{++} and substrate promote a conformation change, repositioning several residues, including the catalytic base. It is speculated that this change in conformation upon substrate binding may add to the substrate selectivity of the enzyme (Gabelli *et al.*, 2002).

It has been demonstrated that during bacterial growth glycogen can be simultaneously synthesized and degraded (Belanger and Hatfull, 1999). The precursor of glycogen, ADP-glucose (ADPG), is synthesized by ADPG pyrophosphorylase. This nucleotide sugar precursor can be degraded by specific pyrophosphatases, such as the Nudix hydrolase family member ASPPase of *E. coli*, encoded by the *aspP* gene. Inactivation of the ASPPase encoding gene produced cells with higher glycogen content than wild-type bacteria, while introduction of an intact copy of the gene in these mutants yielded cells with a dramatically reduced glycogen concentration compared to untransformed mutants. *E. coli* ASPPase was shown to hydrolyze ADP-glucose, ADP-mannose and ADP-ribose (Moreno-Bruna *et al.*, 2001).

The final *E. coli* Nudix hydrolase is *orf135*. Orf135 was found to hydrolyze CTP, dCTP and 5-methyl-dCTP, with the highest catalytic efficiency for the latter substrate (O'Handley *et al.*, 2000). Amongst various oxidized nucleotides tested as substrates for Orf135, only (5-hydroxy-) CTP served as substrates. The catalytic efficiency for 5-hydroxy-CTP hydrolysis was 30 times higher than that for CTP hydrolysis. It has been suggested that 5-OH-

CTP introduces transcriptional errors causing deleterious effects in *E. coli* (Fujikawa and Kasai, 2002).

Nudix hydrolase genes/enzymes in other organisms

The number of Nudix hydrolase encoding genes found in organisms varies greatly from a single one in *Helicobacter pylori* to 30 in *Bacillus anthracis* (Table 1). *Bacillus cereus*, which is closely related to *Bacillus anthracis* (about 90% genomic identity (Read *et al.*, 2003;Xu *et al.*, 2004)), contains 26 Nudix hydrolase encoding genes. If the number of Nudix hydrolase encoding genes is related to the genome size of the organism, *D. radiodurans* contains the highest density of genes encoding this protein family (Table 1). It has been suggested that the high number of Nudix hydrolase encoding genes in *D. radiodurans* (21) may be related to its high degree of radiation resistance. A large number of surveillance enzymes thus may protect this organism from the adverse effects of radiation (Xu *et al.*, 2001). Accordingly, in the gastric pathogen *H. pylori* that lives in a more constant and protected environment, such a high number of surveillance proteins would be redundant. Several Nudix hydrolase proteins of *D. radiodurans* are fused to other domains and form multi-domain proteins, similar to *E. coli* Orf257 and its orthologs in other organisms. The domain composition of some of these proteins indicates that they may be involved in novel DNA-repair pathways. Two *D. radiodurans* Nudix hydrolase proteins contain a duplication of the Nudix hydrolase motif, which has so far not been observed in any other organism (Makarova *et al.*, 2000).

NudA, the only Nudix hydrolase present in *H. pylori*, belongs to the nucleoside polyphosphate hydrolase subgroup. It preferably hydrolyzes Ap_4A to ATP and AMP. Insertional mutagenesis of *nudA* resulted in a 2-7-fold decrease in survival rate compared to the wild type after hydrogen peroxide exposure. Western blot analyses revealed that this protein is expressed constitutively in *H. pylori* at different growth stages and during stress, indicating that the protein has a housecleaning function (Lundin *et al.*, 2003).

The identification of the Nudix hydrolase motif containing ADP-ribose pyrophosphatase from the thermophilic archaeon *Methanococcus jannaschii* indicates that the Nudix hydrolase protein family is represented in all three kingdoms. This highly specific ADP-ribose hydrolase is heat stable up to temperatures of 85^{o}C (Sheikh *et al.*, 1998). *M. jannaschii* does not appear to contain a MutT enzyme, since the ADP-ribose pyrophosphatase is the only Nudix hydrolase predicted from its genome. In obligate anaerobes, oxidation of dGTP may not occur, thus precluding the presence of a MutT enzyme. Similarly, no increase in mutation frequency was observed in an *E. coli* MutT-negative mutant strain grown anaerobically on mineral medium or on rich medium pretreated with an enzyme complex that reduces free oxygen (Fowler *et al.*, 1994). An ADP-ribose pyrophosphatase domain, which shares strong homology with the *M. jannaschii* ADP-ribose pyrophosphatase, is also found in the C-terminal domain of a bifunctional enzyme encoded by the *slr0787* ORF of the cyanobacterium *Synechocystis* sp. The N-terminal domain of this protein is similar to NMN adenylyltransferase of *M. jannaschii* (Raffaelli *et al.*, 1999). NMN adenylyltransferase catalyzes NAD synthesis from NMN and ATP.

Aps1 of *Schizosaccharomyces pombe* (Ingram *et al.*, 1999) and Yor163w of *Saccharomyces cerevisiae* (Cartwright and Lennan, 1999) are two Nudix hydrolases with very similar catalytic properties. Both proteins, which share 43% sequence similarity, show a substrate preference for diadenosine hexaphosphate (Ap_6A), but also Ap_5A, Ap_5 and Ap_4 are hydrolyzed. Distinct from the properties of other Nudix hydrolases catalyzing the hydrolysis of Ap_nA compounds, Yor163w forms a mixture of reaction products. Aps1 hydrolyzes Ap_6A to p_4 and ADP as the major reaction products with no detectable production of p_5A (Ingram *et al.*, 1999). Yor163w, however, hydrolyzes Ap_6A to p_4A + ADP (76%) and p_5A + AMP (24%). The

biological function of both proteins remains to be determined. Neither Ap_6A nor Ap_5A have been reported in any single cell organism, but only in secretory granules of certain specialized mammalian cells. Both p_5A and p_4A have been reported in *S. cerevisiae* during sporulation but not during vegetative growth (Jakubowski, 1986). Because Ap_5A is a potent inhibitor of the essential enzyme adenylate kinase, one possible function of these enzymes may be to eliminate these potentially toxic dinucleotides during sporulation (Cartwright and Lennan, 1999).

Another exceptional Nudix hydrolase found in *S. cerevisiae* is the PCD1 protein. This peroxisomal protein is active toward Coenzyme A and derivatives. Oxidized CoA disulfide is preferred over CoA as a substrate and products are 3'-phosphoadenosine 5'-monophosphate and 4' phosphopantetheine. The authors propose that this enzyme is involved in removal of potentially toxic oxidized CoA disulfide from peroxisomes to maintain the capacity for β-oxidation of fatty acids (Cartwright *et al.*, 2000). A CoA hydrolyzing Nudix hydrolase has also been identified in *D. radiodurans* (Xu *et al.*, 2001). Elucidation of the three dimensional structure of CoA pyrophosphatase of this organism revealed that it contains, in addition to its Nudix hydrolase motif, an N-terminal helix which occupies part of the site required for substrate binding. Thus, substrate binding involves movement of this α-helix from the active site of the enzyme (Kang *et al.*, 2003).

Pur7, a gene product from the puromycin antibiotic biosynthetic gene cluster of *Streptomyces alboniger* (Tercero *et al.*, 1996), is a Nudix hydrolase which clearly fits to the description of being a housecleaning enzyme hydrolyzing potentially toxic compounds (Bessman *et al.*, 1996). Previously, it was thought that Pur7 catalyzed the second step in the biosynthesis of this aminonucleoside antibiotic: the hydrolysis of 3'-keto-3'-dATP resulting in 3'-keto-3'-dAMP and PP_i (Tercero *et al.*, 1996). However, expression and characterization of this protein showed that this intermediate is not a substrate for Pur7. Instead, 3'-amino-3'-dATP and 3'-amino-3'-dTTP were found to be hydrolyzed by Pur7 (with a 30 times lower affinity for 3'-amino-3'-dTTP) (Espinosa *et al.*, 1999). Hydrolysis of 3'-amino-3'-dATP is crucial, since this compound is a potent inhibitor of DNA-dependent RNA polymerase, whereas the reaction products are not (Armstrong and Eckstein, 1976). Thus by removal of the toxic intermediate, Pur7 allows the biosynthetic pathway to continue.

A Nudix hydrolase that has been shown to be involved in biosynthesis is the YlgG of *Lactococcus lactis*. YlgG was shown to exhibit a high affinity Mg^{++}-dependent dihydroneopterin triphosphate (DHNTP) pyrophosphatase activity, which is the second step in the pterin branch of the folate biosynthesis pathway. Inactivation of *ylgG* in *L. lactis* resulted in accumulation of DHNTP and folate depletion (Klaus *et al.*, 2005).

Nudix hydrolase encoding genes are also found in various viral genomes. The first member of the Nudix hydrolase gene superfamily identified in bacterial viruses was the e.1 gene of bacteriophage T4. It was shown to hydrolyze FAD, adenosine 5'-triphospho-5'-adenosine (Ap(3)A), and ADP-ribose. Although the protein was predicted to be orthologous to *E. coli* MutT on the basis of a sequence homology search, the properties of the gene and of the purified protein did not support this notion. Deletion of e.1 did not result in a clear phenotype, since no effect on phage growth and replication could be detected (Xu *et al.*, 2002). The *g5R* gene of the African swine fever virus (ASFV) encodes a Nudix hydrolase preferentially degrading the non-Nudix hydrolase substrate diphosphoinostol polyphosphate, but it is also active with GTP, adenosine 5'-pentaphosphate, dGTP, diadenosine, diguanine nucleotides. Infections with ASFV led to 50% reduction in cellular PP-Ins-P5, ATP and GTP levels. Transient expression of epitope-tagged g5R protein under control of its own promoter in ASFV-infected cells showed an apparent localization in the rough endoplasmatic reticulum. It is suggested that g5R protein regulates a stage of viral morphogenesis, which involves manipulation of components of the cellular secretory pathway by the virus. Diphosphoinositol polyphosphate-mediated membrane

trafficking may thus be disturbed by g5R protein(Cartwright *et al.*, 2002).

Nudix hydrolase motif containing proteins are also encoded by bacterial plasmid DNA. Analysis of the complete nucleotide sequence of the erythromycin resistance plasmid pNG2 of *Corynebacterium diphtheriae* revealed the presence of a Nudix hydrolase gene, of which neither function nor substrate specificity have been studied (Tauch *et al.*, 2003). Also the indigenous plasmid pSAM2 of *Streptomyces ambofaciens* encodes a Nudix hydrolase protein, which was shown to be involved in plasmid immunity of the host cell. Mechanisms of conjugal immunity prevent redundant exchange between two cells harboring the same conjugative plasmid element. The presence of a single copy of the pSAM2 immunity factor gene (*pif*) in the recipient strain was sufficient to abolish both transfer and initiation of plasmid transfer. This is the first protein of the Nudix hydrolase family shown to be involved in bacterial conjugation control (Possoz *et al.*, 2003). Its substrate specificity has not been studied.

Nudix hydrolases hydrolyzing non-Nudix compounds

Diphosphoinositol polyphosphates (PP-$InsP_n$) are known to be involved in some intriguing signal transduction processes (Safrany *et al.*, 1998). These compounds were first characterized in the amoebae *Dictyostelium discoideum* (Stephens *et al.*, 1993), but later they were found to occur in organisms across the phylogenetic spectrum (Safrany *et al.*, 1998). Diphospho-pentakiphosphate (PP-InsP5, Fig. 5) binds to proteins that participate in the control of synaptic vesicle trafficking (Shears *et al.*, 1995). In intact mammalian cells there is an ongoing, rapid metabolic flux through a kinase-phosphatase cycle interconverting $InsP_6$ with PP-$InsP_5$. The level of PP-$InsP_5$ could be decreased following treatment with thapsigargin, a tumor promoter (Safrany *et al.*, 1998). It appears that $InsP_6$ kinase was inhibited by thapsigargin-mediated changes in the status of cellular Ca^{2+} pools (Glennon and Shears, 1993).

The β-phosphate from the diphosphate groups in the most highly phosphorylated members of the inositol-based cell signaling family, PP-$InsP_5$ and $[PP]_2$-$InsP_4$, can be hydrolyzed by an enzyme isolated from rat liver. This Nudix hydrolase motif containing enzyme, diphosphoinositol polyphosphate phosphohydrolase (DIPP), is the first example of a Nudix hydrolase that hydrolyzes a non-Nudix compound. A considerable portion of the monomeric 18 kDa DIPP protein was microsequenced and database searching with the BLAST algorithm revealed a close match between DIPP and a predicted amino acid sequence from a cDNA clone from human uterus. Expression and subsequent purification of the 172 amino acid human DIPP yielded a protein capable of hydrolyzing both PP-$InsP_5$ and $[PP]_2$-$InsP_4$. The Nudix hydrolase domain was shown to be essential in catalysis by site-specific mutagenesis of one of the conserved glutamate residues (E70). Mutant E70Q was inactive with both of the DIPP substrates (Safrany *et al.*, 1998). These findings lead to a revision of the idea that the MutT signature designates a unique binding domain for nucleosides (O'Handley *et al.*, 1996).

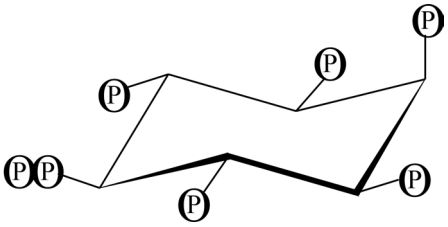

Figure 5. Structure of diphosphopentakiphosphate, the substrate of DIPP.

Part of the human DIPP protein sequence (Val-34 to Glu-85), which includes the MutT motif, is 46% identical to the corresponding regions in Aps1 of *S. pombe* (Ingram *et al.*, 1999) and Yor163w of *S. cerevisiae* (Cartwright and Lennan, 1999). It has been shown that the latter proteins, which are known to hydrolyze Ap_6A and Ap_5A, were also capable to hydrolyze PP-$InsP_5$ and $[PP]_2$-$InsP_4$ and, conversely, that DIPP could hydrolyze Ap_6A and Ap_5A (Safrany *et al.*, 1999). These results provide yet another example of Nudix hydrolases capable to hydrolyze a non-Nudix compound, but also that one protein may catalyze hydrolysis of unrelated classes of substrates. Furthermore, these two groups of metabolites have independently emerged as participants in various aspects of signal transduction (Safrany *et al.*, 1999).

The ability of some Nudix hydrolases to utilize a sugar phosphate as a substrate, prompted testing of 5-phosphoribosyl 1-pyrophosphate (PRPP) as a substrate. PRPP is both a substrate and a regulator of purine, pyrimidine and pyridine nucleotide biosynthesis. A potential product of PRPP pyrophosphatase activity would be ribose 1,5-bisphosphate, which has been shown to be an important regulator of glycolysis by activating phosphofructokinase activity in mammals (Ishikawa *et al.*, 1990). Various dinucleoside polyphosphatases and DIPP enzymes were shown to hydrolyze PRPP, yielding ribose 1,5-bisphosphate as a product. This subfamily of Nudix hydrolase enzymes thus can hydrolyze seemingly unrelated substrates. However, modeling of PRPP onto the crystal structure of *C. elegans* Ap_4A hydrolase (Bailey *et al.*, 2002) shows that it can fit readily in the substrate binding cleft of the enzyme. Whether PRPP hydrolysis is of physiological relevance, remains to be determined. This can be confirmed by measuring PRPP and ribose 1,5-bisphosphate concentrations in cells in which the relevant Nudix hydrolase activities have been eliminated by gene disruption or gene deletion and comparing these with wild type strains (Fisher *et al.*, 2002).

In conclusion

Nudix hydrolase proteins are defined as proteins carrying the conserved $GX_5EX_7REX_2EEXG$ amino acid sequence motif, that can hydrolyze the pyrophosphate bond of specific Nudix (nucleotide diphosphate linked to an indiscriminate moiety X) compound(s). They are ubiquitously spread in nature with representatives in bacteria, archaea and eukaryotic organisms. Genomic data indicates that usually more than one representative of this protein family is present per organism. So far, the analysis of Nudix hydrolase enzymes has mainly focused on determining substrate specificity and catalytic mechanism, with only a few examples of detailed studies on the physiological role of these enzymes in a broader context. Future studies will determine whether members of this protein family can be mainly characterized as "housecleaning" enzymes, or whether the majority of these proteins are involved in cellular or metabolic regulation.