the FEBS Journal

REVIEW ARTICLE

Preventive DNA repair by sanitizing the cellular (deoxy) nucleoside triphosphate pool

Gergely N. Nagy[1,2], Ibolya Leveles[1,2] and Beáta G. Vértessy[1,2]

1 Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary
2 Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Hungary

Keywords
DNA repair, non-canonical nucleotides, NTPases, nucleophilic attack, X-ray crystallography

Correspondence
G. N. Nagy or B. G. Vértessy, Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 2 Magyar Tudósok Körútja, 1117 Budapest, Hungary
Tel: +361 382 6707 (BGV),
+361 382 6762 (GNN)
E-mails: nagy.gergely@ttk.mta.hu; vertessy@mail.bme.hu

(Received 26 April 2014, revised 1 July 2014, accepted 16 July 2014)

doi:10.1111/febs.12941

The occurrence of modified bases in DNA is attributed to some major factors: incorporation of altered nucleotide building blocks and chemical reactions or radiation effects on bases within the DNA structure. Several enzyme families are involved in preventing the incorporation of noncanonical bases playing a 'sanitizing' role. The catalytic mechanism of action of these enzymes has been revealed for a number of representatives in clear structural and kinetic detail. In this review, we focus in detail on those examples where clear evidence has been produced using high-resolution structural studies. Comparing the protein fold and architecture of the enzyme active sites, two main classes of sanitizing deoxyribonucleoside triphosphate pyrophosphatases can be assigned that are distinguished by the site of nucleophilic attack. In enzymes associated with attack at the α-phosphorus, it is shown that coordination of the γ-phosphate group is also ensured by multiple interactions. By contrast, enzymes catalyzing attack at the β-phosphorus atom mainly coordinate the α- and the β-phosphate only. Characteristic differences are also observed with respect to the role of the metal ion cofactor (Mg^{2+}) and the coordination of nucleophilic water. Using different catalytic mechanisms embedded in different protein folds, these enzymes present a clear example of convergent evolution.

Non-orthodox bases in DNA: production and elimination

The chemical space in DNA is considered to be highly limited because only four major bases are responsible for carrying out the primary role of storing information (adenine, thymine, cytosine and guanine). However, this view is rather oversimplified because additional DNA bases may also be present in most organisms; for example, as a result of epigenetic modifications [1,2] or DNA damage [3]. Chemically-modified bases not only serve as recognition foci for DNA damage response processes, but also for signal transduction. Among these pathways, regulation of gene expression in several eukaryotes via 5-methylcytosine present in CpG islands is especially well-known [4], whereas methylated versus non-methylated bases within specific sequence contexts in prokaryotes constitute the signal recognition foci for restriction methylases/endonucleases [5]. Hydrolytic deamination of bases at exocylic amino-positions frequently leads to major perturbation of Watson–Crick base-pairing characteristics: hypoxanthine (deaminated adenine) [6] is a guanine mimic, whereas uracil, a deaminated cytosine, is a thymine mimic [7–9] during DNA synthesis. Similarly, oxidation of bases may also lead to erroneous base pairing, as in the case of 8-oxoguanine, which is

Abbreviations
3D, three-dimensional; dNTP, deoxyribonucleoside triphosphate; dUPNP, 2′-deoxy-uridine-5′-[(α,β)-imido]diphosphate; PDB, Protein Data Bank.

considered to be a frequently occurring oxidation product of oxidative stress that can be mismatched to adenine (resulting in G:C to T:A base pair mutation) [10].

For those chemical modifications that are usually considered to be harmful and hence are the primary foci for initiating DNA repair, reactions leading to the modified base can occur without enzyme catalyzed processes, in a spontaneous, uncontrolled manner [11]. From a chemical point of view, the cellular conditions are highly reactive because water and other reactive metabolites are always present. In aerobic organisms, this situation is even amplified by the ubiquitous occurrence of molecular oxygen and various reactive oxygen species. These conditions lead to numerous potentially harmful chemical reactions involving the nucleic acid base moieties either in the deoxyribonucleoside triphosphate building blocks (dNTPs) or within the DNA context. Because of the less than perfect specificity of most DNA polymerases, modified dNTPs can be easily incorporated into DNA if these still offer base-pairing functional groups and are not overly large. For example, deamination [12], oxo-group formation [13], halogenation [14,15] and alkylation [16] of bases does not prevent these modified bases from ending up in the DNA. Parallel to these two possibilities for the appearance of modified bases in DNA, there are also two general pathways acting against their presence. First, the NTP building blocks containing the modified bases may be hydrolyzed, and hence removed from the nucleotide pool to prevent their incorporation. Second, modified bases within the DNA context are targeted principally by base excision repair [17] and, in some cases, by the mismatch repair machinery [18], or, rarely, by nucleotide excision repair [19] as well. Recently, it has became evident that the cellular 'infrastructure' for base excision and mismatch repair is also exploited in signal transduction leading to immunoglobulin gene diversification and active DNA demethylation [20–24]. Hence, the distinction between the harmful versus signalling role of modified bases becomes less clear and such definitions require an assessment of the cellular context (cell status, developmental stage, cell cycle stage, etc.) as well.

Because the DNA context may easily render some physicochemical protection against unwanted spontaneous reactions, especially within the heterochromatin segments, reactions of the base moieties are more frequent in the nucleotide pool, creating modified dNTPs. Enzymes involved in removing these dNTPs are termed as 'sanitizing' or 'house-cleaning' enzymes [25,26]. Although this term is used for enzymes degrading noncanonical nucleotide precursors for both DNA and RNA or other harmful nucleotide metabolites, we focus here on enzymes performing (d)NTP pool sanitization. Typically, such enzymes are characterized by high-specificity recognition pocket for the modified base, similarly to DNA glycosylases. Several of these enzymes catalyze the cleavage of the phosphoanhydride linkage within the phosphate chain of the NTPs, resulting in a nucleoside monophosphate (nucleotide) and inorganic pyrophosphate, whereas others are reported to possess phosphatase activity instead, such as some representatives of Nudix enzymes [27,28]. This reaction is associated with a large free energy change because the triphosphate chain is degraded, and hence it is strongly exothermic and practically irreversible. This thermodynamic feature not only drives the sanitizing reaction, but also constitutes an energy loss for the cell; hence, the high specificity of the sanitizing enzymes towards the unwanted species with the modified bases is of crucial importance for avoiding wasteful hydrolytic futile cycles. Moreover, in some cases, dNTP sanitizing enzymes also need to possess high specificity towards the sugar moiety to degrade only the deoxyribosenucleotides and not their ribo-counterparts. For example, for dUTPases, the hydrolysis of UTP would be strongly unfavorable because UTP is the building block for RNA synthesis and also contributes to other metabolic pathways, such that these enzymes discriminate not only at the base, but also at the sugar moiety [8,29–31].

To enhance the specificity of sanitizing enzymes and to exclude the unwanted hydrolysis of the canonical NTPs, in addition to specific recognition, a kinetic control is also exploited. The catalytic rate constants of sanitizing enzymes are within the order of 1–20 s^{-1}, and the moderate speed allows the escape of the loosely bound canonical NTPs before they can be hydrolyzed. In many cases, the modest k_{cat} is paired with a low K_M value; hence the catalytic efficiency of these enzymes is considerable [32,33]. Based on *in vitro* kinetic analysis of some of these enzymes, the respective noncanonical nucleotide targets are assumed in some cases to overcompete the hydrolysis of canonical nucleotides, further increasing the selectivity of the enzymes [34,35]. For several such housekeeping enzymes, kinetic and structural biology studies, in parallel with high-resolution three-dimensional (3D) structures, became available in recent years (Table 1). In the present review, we focus on those examples of the sanitizing enzymes where delineation of the reaction pathway has recently been described, based on a combination of the 3D structures determined, in most cases, by X-ray crystallography, as well as kinetic stud-

Table 1. Structures providing insight into the enzyme mechanism of nucleotide pyrophosphatases with housekeeping functions. References to publications providing a structure-based determination of the enzyme mechanism are included. For dUTP at the active site of dimer dUTPase (2CIC), the γ-phosphate moiety was fitted in the electron density at 50% occupancy. In the crystal structure (PDB code: 1EXC) for the Maf enzymes, the ligand dUTP is bound in a catalytically incompetent conformation within the active site: only the triphosphate part of the nucleotide is recognized by the enzyme; other parts are not bound [116]. The YceF-type Maf proteins are active against 7-meGTP but not towards canonical NTPs. The YhdE-type Maf enzymes are active against the conventional nucleotides dTTP, UTP and CTP, as well as against 5-methyl-UTP, pseudo-UTP and 5-methyl-CTP. The K_M for the respective modified nucleotide is one order of magnitude lower than the K_M for conventional nucleotides.

Enzyme	Catalyzed reaction	Position of nucleophilic attack	Number of α-phosphate interactions	Number of β-phosphate interactions	Number of γ-gamma phosphate interactions	PDB code (resolution, where applicable)	Reference
all-β dUTPase	dUTP → dUMP + PP_i	αP	2	4	8	2PY4 (1.49 Å)	[67,97]
ITPase	(d)ITP → (d)IMP + PP_i XTP → XMP + PP_i	αP	6	1	5	2Q16 (1.95 Å)	[35]
Maf	meNTP → meNMP + PP_i	αP	1	1	5	1EXC (2.70 Å)	[75,116]
all-α dUTPase	dUTP → dUMP + PP_i (dUDP → dUMP + P_i)	βP	2	5	1	2CIC (1.70 Å)	[63]
8-oxo-dGTP pyrophosphatase (MutT, hMTH1)	8-oxo-dGTP → 8-oxo-dGMP + PP_i	βP	1	1	1	1TUM (NMR structure) 3A6T 3ZR0	[84,118]
RS21-C6 Mazg-like dCTP pyrophosphatase	d(me)CTP → d(me)CMP + PP_i	βP	1	2	2	2OIG (3.30 Å)	[100,102]

ies in solution. Our main aim is to compare the characteristics of catalytic water and phosphate chain coordination for different groups of sanitizing NTP hydrolyzing enzymes.

Protein families of sanitizing enzymes with structural evidence for the catalytic mechanism

Cleavage of the phosphoanhydride linkage of nucleotides as catalyzed by the protein families of sanitizing enzymes may be initiated by nucleophilic attack at either the α- or the β-phosphorus atom of the triphosphate chain. Groups of these enzymes, belonging to either class, are shown in Fig. 1 together with the protein fold characteristic for each, as well as a schematic representation of the reaction chemistry. Below, we address the different enzyme families following this classification based on the site of the nucleophilic attack. Figure 2 shows the conserved residues and sequence motifs involved in substrate or cofactor accommodation and in the catalytic reaction in the case of each of these protein families. Both the protein fold and the sequence motifs involved in coordination of the substrate, the catalytic water and the cofactor metal ions vary considerably between the families. However, there are also some common characteristics. In all cases where the catalytic water could be located, the residue responsible for its coordination is a conserved aspartate or glutamate and, similarly, conserved aspartates/glutamates are also involved in coordination of the divalent metal ion cofactor. An important difference here is that, in all-β dUTPases, the Mg^{2+} cofactor assists catalysis by determining phosphate chain coordination, whereas, in all-α dUTPase and Nudix nucleotide pyrophosphohydrolases, the cofactors play an additional role in the coordination and deprotonation of the attacking nucleophile [36,37]. Figure 3 shows a panel of the active site close-up of selected representatives for the groups of enzymes listed in Fig. 1. Superposing the active sites, as shown in Fig. 4, shows that, although the base and sugar moieties are differently oriented, the phosphate chain conformation is remarkably similar.

Nucleotide pyrophosphohydrolase enzymes with sanitizing function

Nucleophilic attack on the α-phosphorus

all-β dUTPase

From the structural point of view, all-β dUTPases are characterized by five well-conserved sequence motif (motifs 1–5) (Fig. 2) and an antiparallel β-pleated fold, aptly termed as 'the dUTPase fold' (c.f. Pfam domains) [38]. This secondary and tertiary structural arrangement requires an unexpectedly low level of

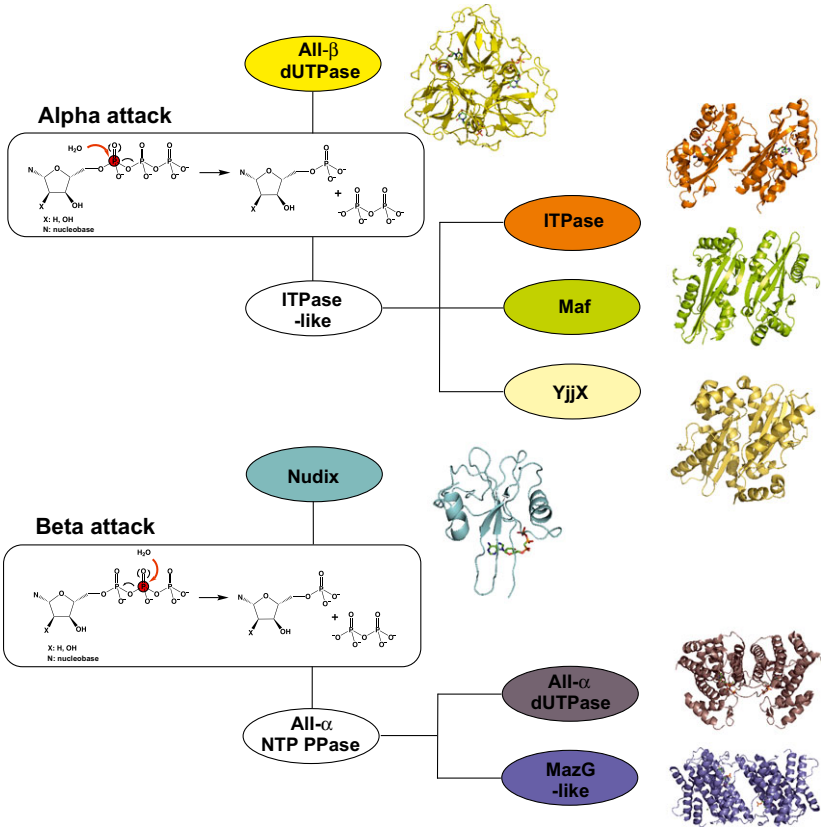

Fig. 1. Catalytic mechanism and evolutionary relationship of housekeeping NTP pyrophosphatases. Housekeeping enzymes are shown with two different catalytic mechanisms leading to nucleotide pyrophosphorolysis; nucleophile attack on α-phosphorus (top), nucleophile attack on β-phosphorus (bottom). The relevant 3D enzyme structures are shown as cartoon models with a stick model of the bound nucleotide ligand (if available). The enzyme structure models selected were (with PDB codes): all-β dUTPase, *Mycobacterium tuberculosis* dUTPase, PDB 2PY4 [31]; ITPase, *E. coli* RdgB ITPase, PDB 2Q16 [35]; Maf, *B. subtilis* Maf, PDB 4HEB [75]; YjjX, PDB 1U5W [74]; Nudix, *E. coli* MutT pyrophosphohydrolase, PDB 1TUM [84]; all-α dUTPase, *C. jejuni* dUTPase, PDB 2CIC [63]; MazG, *D. radiodurans*, MazG-like NTP pyrophosphohydrolase, PDB 2YFD [100]. All-β dUTPases also include monomeric dUTPases [40].

sequence conservation. Although human and Epstein–Barr viral dUTPases show only 19% identity [30,39,40], the same fold is displayed by the protein. The quaternary structure of all-β dUTPases can either be homotrimeric or monomeric. In trimeric enzymes, three actives sites are formed and, in each one of these, amino acids from conserved motifs from all the three subunits participate in substrate binding and/or catalysis [41]. Such intertwined cooperation of three subunits is rather exceptional and directly couples quaternary structure to enzyme activity, as subunits are expected to be inactive on their own [42,43]. In some cases, the trimeric organization is also reflected by the genomic arrangement: in *Caenorhabditis elegans*, the gene of dUTPase contains three slightly different copies in a single reading frame without stop codons, which is expected to result in a 'pseudo-heterotrimeric' protein [44]. Very recently, the same genomic organization was identified in *Drosophila virilis* and investigation of the resulting polypeptide confirmed these expectations (submitted manuscript). Presumably, the joint transcription and translation of the covalently coupled dUTPase 'subunits' may confer some advantage on the phenotype, although these issues need further investigation.

The other group of all-β dUTPases is monomeric, although this notation is somewhat misleading. The genes for these dUTPases, which are present in mammalian herpesviruses but not in the herpesviruses of lower eukaryotes, have evolved by gene duplication where two original dUTPase genes were joined to be transcribed and translated as a covalently coupled 'pseudo-dimeric' dUTPase. Following the presumed gene duplication event, numerous mutations occurred rendering the two copies highly dissimilar. Interestingly, in this protein, the dUTPase fold is still preserved and the active site architecture is also very reminiscent of the situation found in trimeric dUTPases [40,45]. Yet, structural investigation of Epsein–Barr virus dUTPase enabled the hypothesis suggesting a unique conformational dynamics of the C-terminal arm of monomeric dUTPases, with the involvement of a disulfide bond that limits movements of this segment [40].

The mechanism of action of all-β dUTPases has been covered in numerous in-depth studies [8,32,46–55]. The interest in these enzymes is highly motivated on the one hand by the peculiar active site architecture and, on the other hand, by the possibility of using these dUTPases as targets for developing drugs against neoplastic, as well as infectious diseases [31,56–65]. Substrate

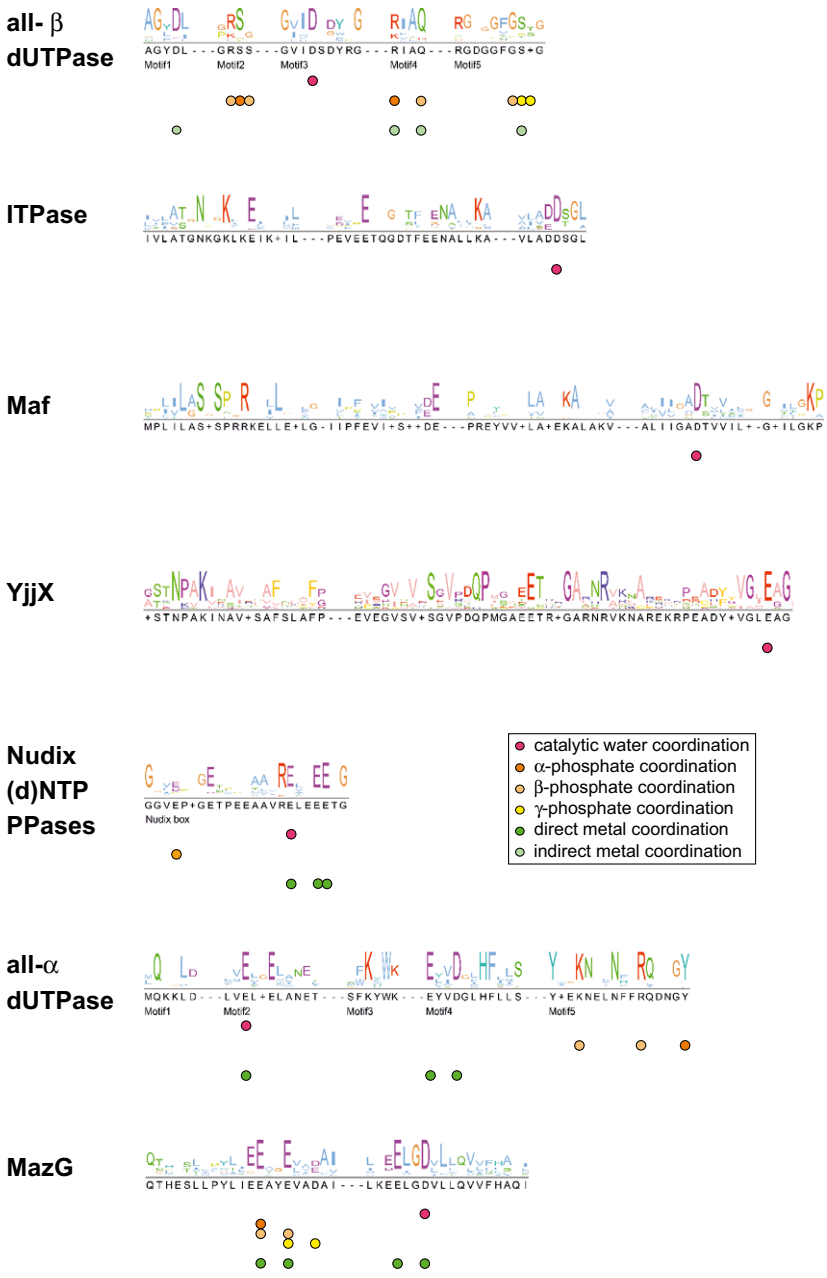

Fig. 2. Amino acid sequence conservation of housekeeping nucleotide pyrophosphatase enzymes. All alignments shown are retrieved from manually curated and high quality *Pfam A* seed alignments [38] using the Pfam accession codes: all-β dUTPase, PF00692; ITPase, PF01725; Maf, PF02545; YjjX, PF01931; Nudix NTPase, PF00293; all-α dUTPase, PF08761; MazG, PF03819. The alignments were manually edited in JALVIEW [117]; a CLUSTALX color code was applied. Sequence conservation of the alignment is displayed with the consensus sequence logo annotation utility built into JALVIEW, with the height of the symbol representing the frequency of residue occurrence. The role of the key conserved residues are depicted as colored circles below the sequence logo using the color code shown. The interactions were identified based on crystal structure with bound NTP ligands, with the exception of YjjX sequence where cognate residues were identified based on sequence similarity to Maf and ITPase enzymes [74]. Nonconserved regions in the sequence are shown with three consecutive spaces. For the sake of clarity, in the case of ITPase, Maf, YjjX and MazG enzymes, only the N-terminal part of the sequence conservation is shown containing the conserved residues that mediate key interactions with the triphosphate part of cognate nucleotides and catalytic divalent metal ions. In the case of all-β and all-α dUTPases, the conserved motifs are displayed as described previously [36,67].

recognition and sharp distinction at both the sugar and at the base level is ensured not only by the side chain of residues but, very importantly, by the protein backbone. For example, the recognition of the substrate within the conserved β-hairpin formed by motif 3 relies heavily on the protein fold and not just on amino acid conservation. This is especially true for the uracil ring, where the specific interactions involve hydrogen bonds between main chain atoms and the uracil functional groups. This has been suggested to potentially resemble an ancient nucleic acid recognition motif [30].

In all cases where data are available, there is clear evidence for a mechanism initiated by nucleophilic attack on the α-phosphorus atom by a water or hydroxide molecule that is accommodated by the strictly conserved aspartate residue in motif 3 (Figs 2 and 3A) [31,39,40,52,66,67]. The other coordinating partners of the nucleophilic catalytic water are usually some other water molecules that are held in place by additional conserved residues. A common characteristic of these dUTPases is that the divalent metal ion cofactor does not make any direct contact to the attacking water.

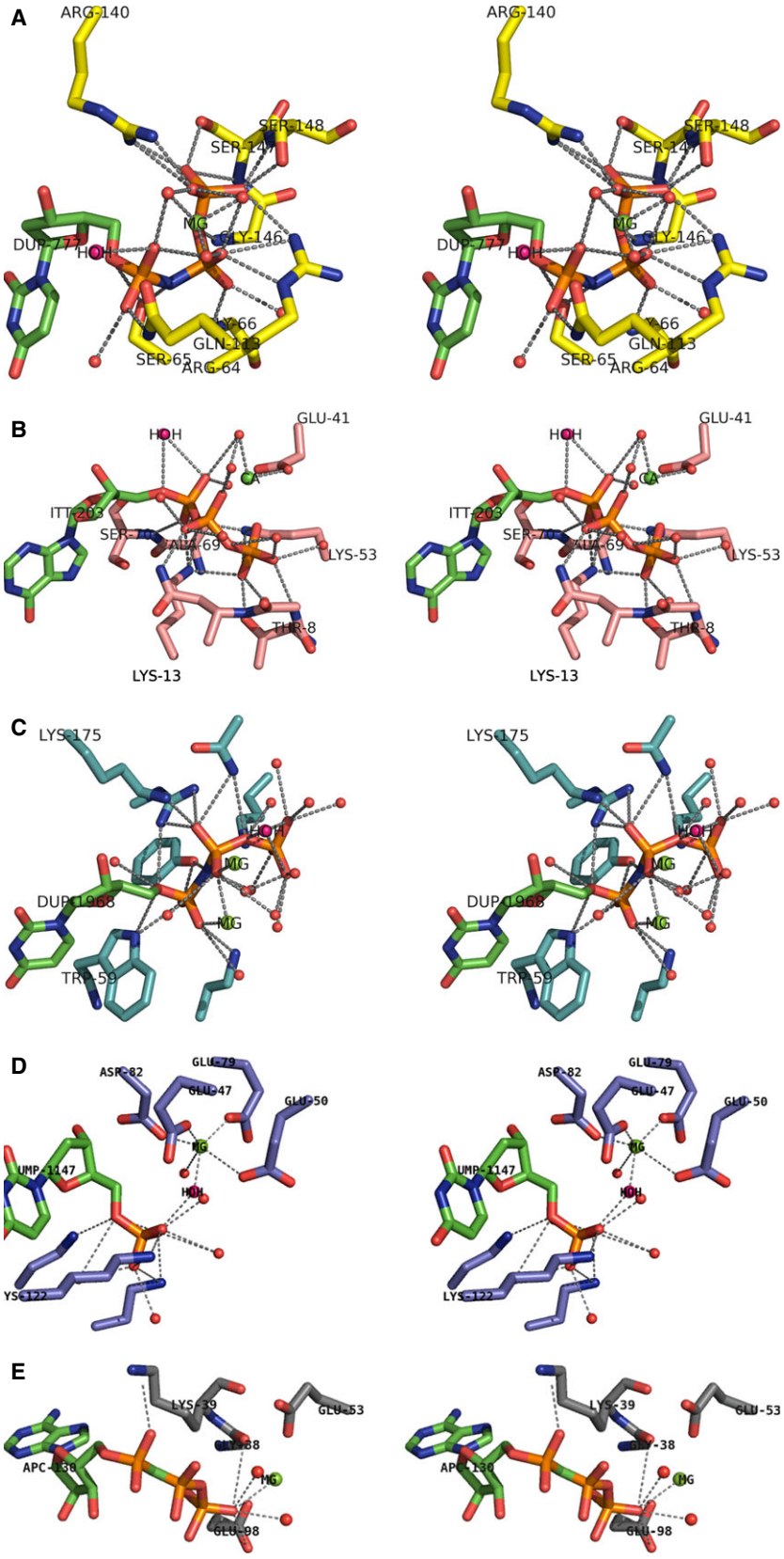

Fig. 3. Close-up to the active site of different dNTPases illuminates triphosphate chain coordination of the respective ligands. (A) all-β dUTPase (PDB code: 2PY4). (B) RdgB ITPase (PDB code: 2Q16). (C) all-α dUTPase (PDB code: 2CIC). (D) MazG (PDB code: 2YFD). (E) MutT (PDB code: 1TUM). Nucleotides and residues, cofactors and solvent interaction partners of their triphosphate chains are shown as ball and stick models with atomic coloring. Polar contacts are marked with dashed lines; metal ion cofactors (Mg^{2+} or Ca^{2+}) cofactor and waters are colored green and red, respectively, whereas catalytic water (as well as an arrow indicating direction of nucleophilic attack) is colored pink. Important residues (three-letter codes), metal ion cofactors (CA, MG) and catalytic waters (HOH) are labelled.

Instead, it is coordinating the phosphate chain of the substrate. Also, interestingly, for several dUTPase from this class, it has been shown that the divalent metal ion is not an absolutely required cofactor; in its absence, k_{cat} is decreased only by a factor of around two, whereas K_M shows an increase of one order of magnitude [51,66,68]. By contrast, in other cases of enzymes using a metal ion cofactor for catalyzing reactions on phosphate groups, the removal of the metal ion renders the enzyme practically inactive. All-β dUTPase also invariably employ a glycine-rich P-loop for catalysis (motif 5), providing critical interactions with the γ-phosphate of the nucleotide substrate. The poor recognition and turnover of dUDP further emphasizes the importance of these interactions [69]. This loop thus stabilizes the catalytically competent phosphate chain conformation involving tridentate coordination of a sole Mg^{2+} cofactor. It also donates an aromatic residue for forming a π-π interaction with the uracil ring [49]. Interestingly, in *Bacillus subtilis* dUTPases, this interaction involves an aromatic ring from another part of the protein sequence, leading to a somewhat different catalytic mechanism [52–54].

Sanitizing enzymes of the ITPase superfamily: ITPase, YjjX and Maf

Defects in purine nucleotide metabolism can result in the incorporation of hypoxanthine and xanthine into DNA and RNA [70]. The respective noncanonical

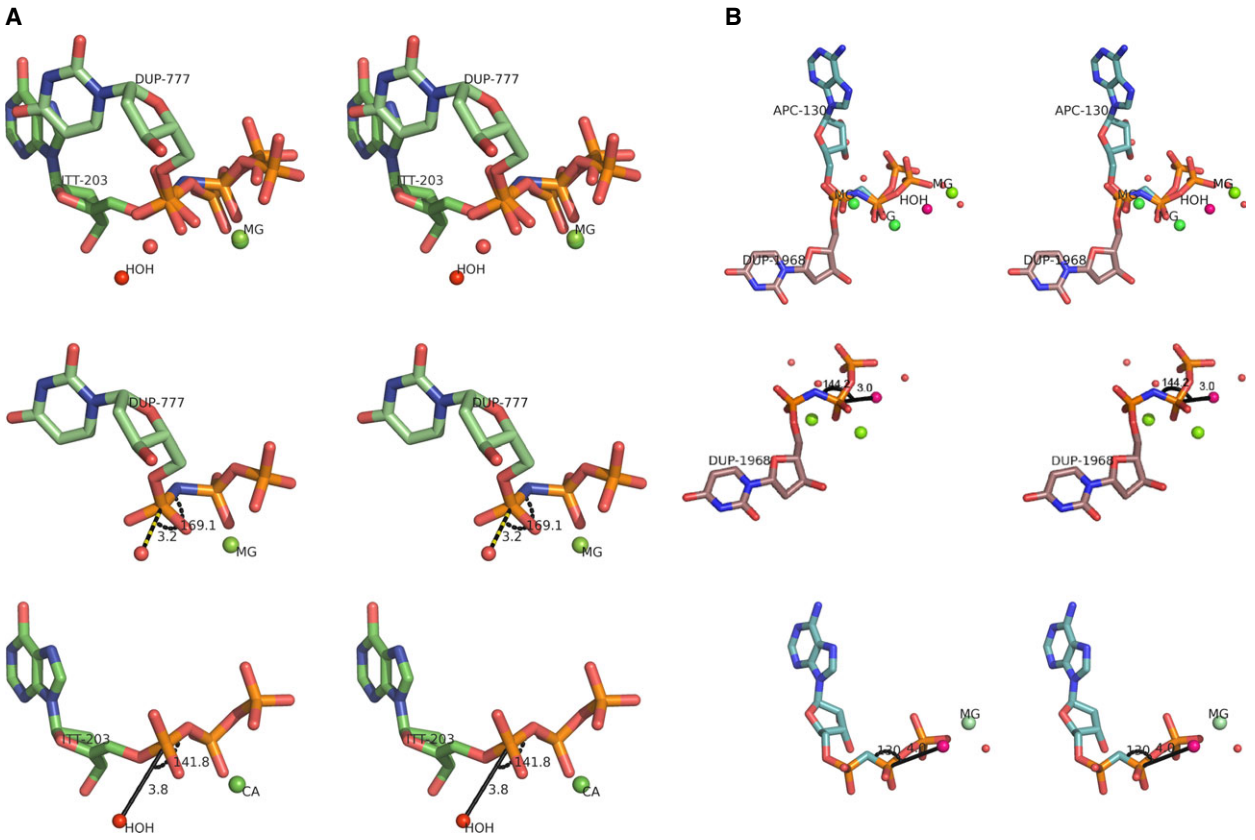

Fig. 4. Orientation of triphosphate chains of respective nucleotide ligands reveal dNTPases with a similar catalytic mechanism. (A) Nucleotides from enzymes effecting hydrolysis with nucleophilic attack on α-phosphorus are superimposed after alignment at the α-phosphate reaction center; dUPNPP from *M. tuberculosis* all-β dUTPase (PDB code: 2PY4), inosine triphosphate from RdgB *E. coli* ITPase-homologue (PDB code: 2Q16). Ball and stick representation with atomic coloring: carbon, green; nitrogen, blue; oxygen, red; phosphorus, orange; Mg^{2+}, light green; catalytic water, pink. Insets also show the geometry of the attacking nucleophile. (B) Nucleotides from enzymes effecting hydrolysis with nucleophilic attack on β-phosphorus are superimposed after alignment at the β-phosphorus reaction center; dUPNPP from *C. jejuni* all-α dUTPase (PDB code: 2CIC), adenosine-5′-[(α,β-methylene)triphosphate] (AMPCPP) from *E. coli* MutT pyrophosphohydrolase (PDB code: 1TUM). Ball and stick representation with atomic coloring: carbon, light blue and faded pink for all-α dUTPase and MutT, respectively; nitrogen, blue; oxygen, red; phosphorus, orange; Mg^{2+}, light green; catalytic nucleophile, pink. Insets also show the geometry of the attacking nucleophile.

deaminated purine NTPs, ITP and XTP, and their reduced derivatives, dITP and dXTP, are degraded by ITPase [71], with homologues HAM1 in yeast [72] and RdgB in *Escherichia coli* [73], together with YjjX [74]. These enzymes belong to the ITPase superfamily, together with Maf enzymes, possessing a broad substrate preference towards both noncanonical (pseudo-UTP, 5-methyl-CTP, 7-methyl-GTP and 5-methyl-UTP) and canonical (dTTP, UTP, CTP) pyrimidine nucleotides [75]. The recently identified HcgB enzyme also belongs here, although it is involved in FeGP cofactor biosynthesis and not in noncanonical nucleotide sanitization [76]. The high sequence similarity of ITPase and Maf suggests their close evolutionary relationship [75], whereas Maf displays a low overall sequential yet high structural similarity to them, with a remarkable correspondance in key conserved active site residues [74], indicating a similar catalytic mechanism (Fig. 2).

The first available 3D structure of the homodimeric ITPase with a bound nucleotide ligand did not provide immediate structure-based explanation for the 100-fold preference of dITP over canonical purine nucleotides GTP or dGTP [73] because, when the *Methanococcus jannaschii* ITPase crystal was soaked with adenosine-5′-[(β,γ)-imido]triphosphate (AMPPNP) nucleotide analogue, the purine ring of the ligand pointed away from the protein, without any specific interaction [71]. However, structure with bound own ligand ITP (Protein Data Bank (PDB) code: 2J4E) revealed the true binding site for the physiological substrate, leaving the cavity associated with AMPPNP coordination as a secondary or artificial site [77]. Binding of ITP to the active site of the homodimer enzyme appeared to utilize an induced fit mechanism, including rigid body displacement leading to the closure of active site, as judged by a comparison of nucleotide bound and free structures [35,78,79]. Mg^{2+} was also observed in proximity to the β- and γ-phosphates, directly coordinated by a Glu residue, which is proposed to be the position of the physiological divalent metal cofactor. The presence of a single metal ion in the active site of the above ITPase, together with structural evidence from other ITPase orthologues, indicated that a single divalent cofactor is sufficient for ITPase catalysis [79]. This structure also revealed specific recognition of the 2-keto oxygen of ITP, with a hydrogen bond and explained the enzyme reactivity differences between inosine and guanosine nucleotides. The role of the major residue determinants that were considered to be responsible for the specific recognition or steric exclusion of the respective purine bases was later also confirmed by site-directed mutagenesis [80]. A model for the catalytic mechanism was provided based on structural insights from the ITPase homologue RgdB [35]. The role of a conserved Asp as a general base, coordinating and deprotonating the catalytic nucleophile for in-line nucleophilic attack on the α-phosphate, was confirmed by structural analyses (PDB code: 2Q16) (Figs 3B and 4A), as well as by site-directed mutagenesis [35,78,80]. Post-catalytic deprotonation of the general base after completion of hydrolysis was suggested to be performed by the nearby Lys53. A key role was also indicated for Lys13 that contacts the bridging oxygen between α-P and β-P and was suggested to enchance polarization of this scissile bond, thereby stabilizing the developing negative charge on that oxygen in the transition state. This lysine, together with the general base Asp and the cofactor-binding Glu, was suggested to form a catalytic triad within the active site. Subsequently, it was shown that simultaneous accommodation of IMP and a sulfate (or alternatively, a phosphate) anion oriented the ITPase active site to a closed position [81], highlighting the critical role of γ-phosphate coordination in ligand-induced active site closure.

Nucleophilic attack on the β-phosphorus

Nudix dNTP pyrophosphatases

Nudix enzymes, characterized by the conserved Nudix domain $GX_5EX_7REUXEEXGU$ (where U represents hydrophobic residues such as Ile, Val and Leu) constitute a large enzyme superfamily [82]. They were initially associated with the sanitization of potentially toxic compounds; subsequently, their participation in various signalling and metabolic pathways has been revealed. To fulfil their diverse functions, these enzymes employ various substrates yet, in this review, we focus only on enzymes performing (d)NTP conversion. MutT from *E. coli* was identified as the first Nudix enzyme with a key role in maintaining both DNA replication and translational fidelity by performing pyrophosphorolysis on oxidized nucleotides 8-oxo-dGTP and 8-oxo GTP [83]. The catalytic mechanism of this enzyme was assessed firstly by the determination of its solution structure in complex with AMPCPP substrate analogue [84], and secondly, by in-depth kinetic characterization [85]. Briefly, similar to all other Nudix enzymes, MutT shows an absolute requirement for divalent cations. This is correlated with the high conservation of acidic residues within the Nudix box, which is responsible for positioning these ions at the active site. The two coordinated metal ions in turn position the catalytic nucleophile for an in-line attack

on the β-phosphorus of the substrate. An additional key role was determined for a Glu catalytic base, which provides substantial rate acceleration as a result of nucleophile deprotonation and the Lys and Arg residues involved in substrate and catalytic base coordination. Although MutT exhibits high substrate specificity for the hydrolysis of 8-oxo dGTP [86], its human homologue, hMTH1, sanitizes numerous other oxidized purine nucleotides, such as 2-OH-dATP, 2-OH-ATP and 8-OH-dATP, and has a much lower preference for the hydrolysis of 8-oxo dGTP over its natural nucleotide counterpart dGTP [87].

Dimeric dUTPase

A group of dUTPases has been shown to form homodimers in solution and possess sequence conservation that is different from the monomeric or trimeric dUTPases [88]. These results, combined with phylogenetic analysis, allowed the definition of a new superfamily of d(C/U)TPases [36], including several distinct enzyme families (dUTPases in trypanosomatides, *Campylobacter jejuni* and dCTP/dUTPases in some Gram-negative bacteria and T4-like phages, as well as dUTPases in various Gram-positive bacteria and their phages) [89–91]. This branch of dUTPase enzymes was later confirmed to belong to the all-α NTP pyrophosphohydrolase superfamily, possessing house-cleaning functions [92].

Kinetic characterization of these novel forms of dUTPases showed that they accepted either dUDP or dUTP as a substrate, whereas the dUMP product was found to be their potent inhibitor, which clearly indicated the distinct mode of action of these enzymes [88,93]. The first dimer dUTPase structure (*Trypanosoma cruzi* dUTPase) revealed many structural and mechanistic aspects [94]. The predominantly helical structure contains two distinct domains with a rigid and mobile character. The rigid domains create the dimer interface where the two active sites of the dimers are located. The mobile domains go through a series of structural rearrangements upon substrate binding, with a loop covering the active site in a stable and tight conformation. A catalytic Mg^{2+} binding site was proposed based on the negative potential at the active site induced by acidic residues; however, the metal cofactor was omitted from the crystallization conditions. Phosphates of bound dUDP were found in a staggered conformation, in contrast to the eclipsed conformation found for trimeric dUTPases [95]. The deoxyuridine part of the substrate adopted a significantly different conformation from that observed in all-β dUTPases.

The structure of *C. jejuni* dUTPase enabled a more precise understanding of the mechanism of action of dimeric dUTPases [36]. This 3D structure displayed the accommodation of 2′-deoxy-uridine-5′-[(α,β)-imido] diphosphate (dUPNP) substrate analogue at the active site, together with physiologically relevant metal ions, coordinated by three conserved glutamate residues and one aspartate residue. In addition, the catalytic nucleophile was clearly indentified within the active site. The proposed water molecule is coordinated by two Mg^{2+} ions and is geometrically well positioned for an in-line attack on the β-phosphate, being only 17° off the direct line expected from an associative S_N2-type mechanism, at a distance of 3.2 Å from the β-phosphorus (Fig. 3C). Sequence and structural comparisons of homologous proteins revealed a high degree of conservation for most of the ligand binding residues. This enabled the formation of a conserved uracil binding site that has a critical role in nucleobase discrimination.

A further study investigated how substrate binding promoted an induced fit that creates the pre-catalytic conformation of dimeric dUTPases [63]. To reveal the molecular determinants of this motion, dU and dUMP binding was assessed in solution by tryptophan fluorescence and in a crystalline state. Both approaches indicated that the presence of pyrophosphate or phosphate mimics substantially alleviated binding of the uridine molecule and the concomitant active site closure, which emphasizes the critical role of α-phosphate coordination. Furthermore, for the first time, a dimeric dUTPase structure with bound NTP analogue was reported (PDB code: 2CIC) [63]. The fact that two instead of three Mg^{2+} ions are accommodated at the active site, compared to the structure with the nucleotide diphosphate analogue (PDB code: 1W2Y) [36], was explained by the differing cofactor involvement in the catalytic mechanism when converting its dUTP or dUDP substrates. The γ-phosphate of the substrate was only loosely coordinated, with very few interactions observed, and could be fitted to the electron density only with much higher crystallographic B-factors than the α- or β-phosphates upon structural refinement. Subsequent study of typanosomatyde dUTPases provided a solid basis for the confirmation of the mechanism of dimeric dUTPases featuring nucleophilic attack on β-phosphorus [96]. ^{31}P NMR studies performed in $H_2{}^{18}O$ enriched water reported ^{18}O incorporation into PP_i upon dUTP hydrolysis, in accordance with the similar approach used for all-α dUTPase [97]. The mechanism was also confirmed by an intrinsic tryptophan fluorescence experiment, with the involvement of known transition state mimics, AlF_3 and

MgF_3^-. In contrast to the addition of dU + AlF_3, only dUMP + AlF_3 co-addition exhibited an increased affinity for the enzyme; thus, this complex, mimicking nucleotide diphosphate, was shown to be the more authentic analogue of the transition state of the enzyme. Crystallization of the enzyme with the MgF_3^- and AlF_3 provided further valuable illustration of the transition state structure of the enzyme [96]. This well-described catalytic mechanism may have relevance for other members of the all-α NTP pyrophosphohydrolase enzyme family.

MazG

The MazG domain-containing enzymes, widespread among bacteria, eukaryotes and Archaea, were initially identified as highly conserved NTP pyrophosphatases that are able to hydrolyze all eight canonical nucleotides, albeit with low specificity [98]. The first 3D structure of a MazG enzyme from *Sulfolobus solfataricus* (PDB code: 1VMG) enabled its classification into the newly-identified all-α NTP PPase superfamily by sequence- and structure-based comparisons [92]. The cognate ligand, found with poor resolution at the active sites of the tetramer structure was identified as 2-oxo dATP. Although this was proposed to be the physiological substrate of the enzyme, there was an initial challange in identification of the natural substrate for MazG enzymes [26]. Subsequently, in addition to the hydrolysis of canonical (d)NTPs, a rather diverse substrate specificity on noncanonical nucleotides was reported by the characterization of various MazG homologues. The biological role of tandem MazG domain proteins was reported to be the regulation of ppGpp (guanosine 3′,5′-bispyrophosphate) cellular levels, extending the period of bacterial cell survival under nutritional stress [99,100]. MazG from *Deinococcus radiodurans* displayed not only structural, but also functional similarity to dimeric dUTPases by its high specificity to dUTP along with the inability to hydrolyze dTTP, leading to the hypothesis of a MazG-like ancestor as the evolutionary precursor of dimeric dUTPases [100]. The RS21-C6 mammalian MazG displayed preference for 5-halo-dCTP analogues [101], whereas the MazG from *Mycobacteria* was shown to be responsible for the degradation of 5-OH-dCTP [99]. Numerous structural and mutational studies have shed light on the active site architecture and mechanism of these enzymes [100,102,103]. In the MazG enzymes, the E*EXX*E_{12-28}E*XX*D conserved sequence motif positions Mg^{2+} in the active site, identically to other enzyme representatives from the all-α NTPases superfamily (Fig. 2). Comparing open and closed active site conformations in the apo and dUMP hydrolysis product complexed states revealed a MazG-specific latch that goes through major conformational changes upon forming contact with the deoxyribose and α-phosphate moieties of the bound nucleotide. In addition, a water bound to Asp of the E*XX*D motif was identified as the catalytic nucleophile [100] based on its similar position to the catalytic water identified in *C. jejuni* dUTPase [36] (Fig. 3C,D).

Enzymes with different fold employ different catalytic mechanism to perform similar function

Having reviewed the biochemical, kinetic and structural insights obtained from the noncanonical (d)NTP sanitizing enzymes, it is apparent that they are involved in (d)NMP formation, in most cases by performing pyrophosphorolysis. The production of the monophosphate, as opposed to the diphosphate, is likely to be important for preventing regeneration of the congate NTP by the non-base specific nucleotide diphosphate kinase [104]. By contrast, base-specific (deoxy)nucleoside monophosphate kinases are employed for (d)NDP generation, which circumvents the effective recycling of modified nucleotides. To our knowledge, this may apply to all of the nucleotide sanitizing enzymes and hence implicates a direct link between their biochemical role and the chemical mechanism. The functional role of (d)NTP PPases fulfils the employment of two main types of catalytic mechanisms promoting different reaction chemistry. Studies on all-β dUTPases present potentially the most in-depth enzyme kinetic and structure-based evidence for the mechanism of action leading to nucleophilic attack on nucleotide α-phosphate. The key features of this mechanism include (a) a conserved acidic residue (Asp) coordinating the nucleophile in an in-line position to α-phosphorus and deprotonating it before attack; (b) triphosphate chain positioned for catalysis by a single Mg^{2+} cofactor; (c) charge-stabilization of the bridging oxygen between α- and β-phosphate by a conserved Ser residue; and (d) tight γ-phosphate coordination upon catalysis performed by a mobile glycine-rich P-loop [8,29,30,32, 39,51–53,65,67,69]. The catalytic mechanism of ITPases exhibits high similarities in all of these structural solutions, with a Lys performing the role described previously [35] and providing an alternative solution for γ-phosphate coordination compared to the P-loop motif [35,79,81]. Less evidence is provided for the catalytic mechanism of Maf enzymes; nevertheless, because these are evolutionary and structural homologues of ITPases, a similar mechanism was proposed for them

[35]. Indeed, analysis of the recently published Maf structures identified the same set of key conserved residues as in the case of ITPases (Fig. 2): an Asp as general base, a Lys that stabilizes the protonated form of the former and another Lys that assists the transition state by charged stabilization [75].

By contrast, the catalytic mechanism of dimeric dUTPases is governed by multiple metal cofactors that play determinant roles in both the coordination of the triphosphate chain and nucleophile catalysis, directly coordinating the nucleophile and assisting with the departure of the leaving group. This was predicted [36] to be similar to that deciphered for DNA polymerases in clear-cut structural detail [105]. The conserved general base Asp or Glu positions the nucleophile for an in-line attack on the β-phosphate. A cluster of conserved acidic residues is responsible for coordination of the two or three Mg^{2+} required for catalysis. Importantly, although the α- and β-phosphates of the nucleotides are coordinated with multiple residue interactions, the γ-phosphate position is less fixed. This mechanism is reminiscent of that determined for Nudix (d)NTP pyrophosphatases [84,85] (Fig. 3E; for a superposition of the respective NTPs, see Fig. 4B).

Within the enzymes utilizing the two above discussed catalytic mechanisms, all-α (dimeric) and all-β (monomeric or trimeric) dUTPases are peculiar representatives satisfying the same enzymatic function [26]. Both dUTPase enzymes display a kinetic trend for a low k_{cat} paired with a low K_M, suggesting that high specificity is their key functional role [8]. These enzymes provide very good additional evidence for the phenomenon of functional convergence in enzyme evolution having emerged in the form of analogous enzymes with recurrence of the same enzyme activity within different structural scaffolds [106].

Functional role of (d)NTP pyrophosphatase enzymes in preventive repair and drug development aspects

Nucleotide pool sanitizing enzymes are working hand in hand together with DNA repair enzymes, such as the 8-oxo-G repair enzymes (MutY, MutM, MutT) [107], and have key role in maintaining genomic integrity. In addition, many NTP pyrophosphatase enzymes were shown to be involved in both noncanonical nucleotide sanitization and canonical (d)NTP pool balance, either directly through catalytic conversion (dCTPP1, belonging to MazG protein family [75,108] or indirectly through nucleotide salvage pathways, such as dUTPase [8] and ITPase [109].

The importance of these enzymes becomes apparent when considering naturally occurring enzyme deficiencies and single nucleotide polimorphisms, as well as induced gene silencing or knockout studies. In some cases, sanitizing enzymes of different origin display overlapping substrate preferences. dUTP elimination is performed by a plethora of enzymes, including MazG RS21-C6 and DR2231 [100] enzymes, as well as isoforms of all-α and all-β dUTPases.

The ubiquity of the *dut* gene in Eukarya, Eubacteria and Archaea implies its existence in the last common ancestor of the three domains of life. The contribution of dUTP hydrolysis to preventive DNA repair, as well as dTTP synthesis, by supplying dUMP as a precursor, indicates that this pathway may have emerged during the advent of thymine instead of uracil incorporation in the nucleic acid carrying the genetic code.

Differences of dUTPase enzymes from parasites and the human host were suggested to contribute to different drug–enzyme recognition characteristics that can be exploited in the design of drugs against protozoan infections [64]. Human dUTPase is also an important target for chemotherapeutic agents that combat a variety of neoplastic diseases [58,110]. The discovery of a novel class of potent human dUTPase inhibitors remarkably enhanced the antitumor activity of thymidylate synthase inhibitors and the efficacy of 5-fluorouracil-based chemotherapy.

The incorporation of dITP or dXTP into DNA is not mutagenic in itself. However, as the DNA repair machinery attempts its removal, this can lead to DNA strand breaks [109]: the biological significance of ITPase is indicated by the lethal effect of the knockout (–/– genotype) of the respective gene encoding ITPase in mice [111], as well as the reported adverse drug reactions associated with a common human ITPase polymorph allele 94C/A (Pro32T) [112]. The ITPA variant in humans is associated with drug sensitivity [112], as well as with elevated level of DNA strand breaks.

Catalytic activity of the Nudix MTH1, the normally non-essential human orthologue of MutT, was demonstrated to be required for cancer cell survival through sanitization of the high level of oxidized DNA metabolites [113]. Therefore, this enzyme may represent a new attractive target strategy for difficult-to-treat tumours that display high levels of replicative and oxidative stress. It has been shown that structurally different selective inhibitors of MTH1 bind to the enzyme via hydrogen-bond interactions with the base moiety recognition pocket, ensuring selectivity over other NTP PPases [114]. A selective inhibitor of MTH1 induced an increase in DNA single-strand breaks, activated DNA repair in human colon carcinoma cells

and effectively suppressed tumour growth in animal models. It has been further reinforced that these recent studies resolve the apparent paradox that inhibition of an anti-oxidant enzyme suppresses tumorigenesis, which may in turn result in a change to the present dietary recommendation of antioxidants for cancer prevention [115].

Acknowledgements

This work was supported by the Hungarian Scientific Research Fund OTKA NK 84008 and K109486, the Baross Program of the New Hungary Development Plan (3DSTRUCT, OMFB-00266/2010 REG-KM-09-1-2009-0050), the Hungarian Academy of Sciences (TTK IF-28/2012 and the MedinProt program) and the European Commission FP7 Biostruct-X project (contract no. 283570).

Author contribution

The review was written by G. N. Nagy and G. B. Vertessy. Figures were created by G. N. Nagy and I. Leveles.

References

1 Kriukiene E, Liutkeviciute Z & Klimasauskas S (2012) 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chem Soc Rev* **41**, 6916–6930.
2 Weber M & Schubeler D (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**, 273–280.
3 Basu A, Broyde S, Iwai S & Kisker C (2010) DNA damage, mutagenesis, and DNA repair. *J Nucleic Acids* **2010**, 182894.
4 Cooper DN, Mort M, Stenson PD, Ball EV & Chuzhanova NA (2010) Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum Genomics* **4**, 406–410.
5 Marinus MG & Casadesus J (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**, 488–503.
6 Parry TE (2007) On the mutagenic action of adenine. *Leuk Res* **31**, 1621–1624.
7 Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.
8 Vertessy BG & Toth J (2009) Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res* **42**, 97–106.
9 Lari SU, Chen CY, Vertessy BG, Morre J & Bennett SE (2006) Quantitative determination of uracil residues in *Escherichia coli* DNA: contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst)* **5**, 1407–1420.
10 Cheng KC, Cahill DS, Kasai H, Nishimura S & Loeb LA (1992) 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G—T and A—C substitutions. *J Biol Chem* **267**, 166–172.
11 Jackson AL & Loeb LA (2001) The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat Res* **477**, 7–21.
12 Kow YW (2002) Repair of deaminated bases in DNA. *Free Radic Biol Med* **33**, 886–893.
13 Ventura I, Russo MT, De Luca G & Bignami M (2010) Oxidized purine nucleotides, genome instability and neurodegeneration. *Mutat Res* **703**, 59–65.
14 Kawai Y, Morinaga H, Kondo H, Miyoshi N, Nakamura Y, Uchida K & Osawa T (2004) Endogenous formation of novel halogenated 2′-deoxycytidine. Hypohalous acid-mediated DNA modification at the site of inflammation. *J Biol Chem* **279**, 51241–51249.
15 Pettersen HS, Visnes T, Vagbo CB, Svaasand EK, Doseth B, Slupphaug G, Kavli B & Krokan HE (2011) UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. *Nucleic Acids Res* **39**, 8430–8444.
16 Sedgwick B, Bates PA, Paik J, Jacobs SC & Lindahl T (2007) Repair of alkylated DNA: recent advances. *DNA Repair (Amst)* **6**, 429–442.
17 Krokan HE & Bjoras M (2013) Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583.
18 Russo MT, De Luca G, Degan P & Bignami M (2007) Different DNA repair strategies to combat the threat from 8-oxoguanine. *Mutat Res* **614**, 69–76.
19 Boiteux S, Gellon L & Guibourt N (2002) Repair of 8-oxoguanine in *Saccharomyces cerevisiae*: interplay of DNA repair and replication mechanisms. *Free Radic Biol Med* **32**, 1244–1253.
20 Maul RW, Saribasak H, Martomo SA, McClure RL, Yang W, Vaisman A, Gramlich HS, Schatz DG, Woodgate R, Wilson DM 3rd *et al.* (2011) Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol* **12**, 70–76.
21 Maul RW & Gearhart PJ (2010) AID and somatic hypermutation. *Adv Immunol* **105**, 159–191.
22 Gazumyan A, Bothmer A, Klein IA, Nussenzweig MC & McBride KM (2012) Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. *Adv Cancer Res* **113**, 167–190.
23 Wu H & Zhang Y (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68.

24 Piccolo FM & Fisher AG (2013) Getting rid of DNA methylation. *Trends Cell Biol* **24**, 136–143.
25 Bessman MJ, Frick DN & O'Handley SF (1996) The MutT proteins or 'Nudix' hydrolases, a family of versatile, widely distributed, 'housecleaning' enzymes. *J Biol Chem* **271**, 25059–25062.
26 Galperin MY, Moroz OV, Wilson KS & Murzin AG (2006) House cleaning, a part of good housekeeping. *Mol Microbiol* **59**, 5–19.
27 Kawasaki K, Kanaba T, Yoneyama M, Murata-Kamiya N, Kojima C, Ito Y, Kamiya H & Mishima M (2012) Insights into substrate recognition by the *Escherichia coli* Orf135 protein through its solution structure. *Biochem Biophys Res Commun* **420**, 263–268.
28 Patil AG, Sang PB, Govindan A & Varshney U (2013) *Mycobacterium tuberculosis* MutT1 (Rv2985) and ADPRase (Rv1700) proteins constitute a two-stage mechanism of 8-oxo-dGTP and 8-oxo-GTP detoxification and adenosine to cytidine mutation avoidance. *J Biol Chem* **288**, 11252–11262.
29 Larsson G, Svensson LA & Nyman PO (1996) Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analogue (dUDP). *Nat Struct Biol* **3**, 532–538.
30 Mol CD, Harris JM, McIntosh EM & Tainer JA (1996) Human dUTP pyrophosphatase: uracil recognition by a beta hairpin and active sites formed by three separate subunits. *Structure* **4**, 1077–1092.
31 Varga B, Barabas O, Takacs E, Nagy N, Nagy P & Vertessy BG (2008) Active site of mycobacterial dUTPase: structural characteristics and a built-in sensor. *Biochem Biophys Res Commun* **373**, 8–13.
32 Larsson G, Nyman PO & Kvassman JO (1996) Kinetic characterization of dUTPase from *Escherichia coli*. *J Biol Chem* **271**, 24010–24016.
33 Toth J, Varga B, Kovacs M, Malnasi-Csizmadia A & Vertessy BG (2007) Kinetic mechanism of human dUTPase, an essential nucleotide pyrophosphatase enzyme. *J Biol Chem* **282**, 33572–33582.
34 Saraswat V, Massiah MA, Lopez G, Amzel LM & Mildvan AS (2002) Interactions of the products, 8-oxo-dGMP, dGMP, and pyrophosphate with the MutT nucleoside triphosphate pyrophosphohydrolase. *Biochemistry* **41**, 15566–15577.
35 Savchenko A, Proudfoot M, Skarina T, Singer A, Litvinova O, Sanishvili R, Brown G, Chirgadze N & Yakunin AF (2007) Molecular basis of the antimutagenic activity of the house-cleaning inosine triphosphate pyrophosphatase RdgB from *Escherichia coli*. *J Mol Biol* **374**, 1091–1103.
36 Moroz OV, Harkiolaki M, Galperin MY, Vagin AA, Gonzalez-Pacanowska D & Wilson KS (2004) The crystal structure of a complex of *Campylobacter jejuni* dUTPase with substrate analogue sheds light on the mechanism and suggests the 'basic module' for dimeric d(C/U)TPases. *J Mol Biol* **342**, 1583–1597.
37 Mildvan AS, Xia Z, Azurmendi HF, Saraswat V, Legler PM, Massiah MA, Gabelli SB, Bianchet MA, Kang LW & Amzel LM (2005) Structures and mechanisms of Nudix hydrolases. *Arch Biochem Biophys* **433**, 129–143.
38 Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res* **40**, D290–D301.
39 Varga B, Barabas O, Kovari J, Toth J, Hunyadi-Gulyas E, Klement E, Medzihradszky KF, Tolgyesi F, Fidy J & Vertessy BG (2007) Active site closure facilitates juxtaposition of reactant atoms for initiation of catalysis by human dUTPase. *FEBS Lett* **581**, 4783–4788.
40 Tarbouriech N, Buisson M, Seigneurin JM, Cusack S & Burmeister WP (2005) The monomeric dUTPase from Epstein-Barr virus mimics trimeric dUTPases. *Structure* **13**, 1299–1310.
41 Fiser A & Vertessy BG (2000) Altered subunit communication in subfamilies of trimeric dUTPases. *Biochem Biophys Res Commun* **279**, 534–542.
42 Takacs E, Grolmusz VK & Vertessy BG (2004) A tradeoff between protein stability and conformational mobility in homotrimeric dUTPases. *FEBS Lett* **566**, 48–54.
43 Kovari J, Barabas O, Takacs E, Bekesi A, Dubrovay Z, Pongracz V, Zagyva I, Imre T, Szabo P & Vertessy BG (2004) Altered active site flexibility and a structural metal-binding site in eukaryotic dUTPase: kinetic characterization, folding, and crystallographic studies of the homotrimeric Drosophila enzyme. *J Biol Chem* **279**, 17932–17944.
44 McClure MA (2001) Evolution of the DUT gene: horizontal transfer between host and pathogen in all three domains of life. *Curr Protein Pept Sci* **2**, 313–324.
45 McGeoch DJ (1990) Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res* **18**, 4105–4110.
46 Nord J, Nyman P, Larsson G & Drakenberg T (2001) The C-terminus of dUTPase: observation on flexibility using NMR. *FEBS Lett* **492**, 228–232.
47 Vertessy BG, Larsson G, Persson T, Bergman AC, Persson R & Nyman PO (1998) The complete triphosphate moiety of non-hydrolyzable substrate analogues is required for a conformational shift of the flexible C-terminus in *E. coli* dUTP pyrophosphatase. *FEBS Lett* **421**, 83–88.
48 Palmen LG, Becker K, Bulow L & Kvassman JO (2008) A double role for a strictly conserved serine:

further insights into the dUTPase catalytic mechanism. *Biochemistry* **47**, 7863–7874.
49 Pecsi I, Leveles I, Harmat V, Vertessy BG & Toth J (2010) Aromatic stacking between nucleobase and enzyme promotes phosphate ester hydrolysis in dUTPase. *Nucleic Acids Res* **38**, 7179–7186.
50 Takacs E, Barabas O, Petoukhov MV, Svergun DI & Vertessy BG (2009) Molecular shape and prominent role of beta-strand swapping in organization of dUTPase oligomers. *FEBS Lett* **583**, 865–871.
51 Kovari J, Barabas O, Varga B, Bekesi A, Tolgyesi F, Fidy J, Nagy J & Vertessy BG (2008) Methylene substitution at the alpha-beta bridging position within the phosphate chain of dUDP profoundly perturbs ligand accommodation into the dUTPase active site. *Proteins* **71**, 308–319.
52 García-Nafría J, Timm J, Harrison C, Turkenburg JP & Wilson KS (2013) Tying down the arm in Bacillus dUTPase: structure and mechanism. *Acta Cryst D* **D69**, 1367–1380.
53 Garcia-Nafria J, Harkiolaki M, Persson R, Fogg MJ & Wilson KS (2011) The structure of *Bacillus subtilis* SPbeta prophage dUTPase and its complexes with two nucleotides. *Acta Crystallogr D Biol Crystallogr* **67**, 167–175.
54 Garcia-Nafria J, Burchell L, Takezawa M, Rzechorzek NJ, Fogg MJ & Wilson KS (2010) The structure of the genomic *Bacillus subtilis* dUTPase: novel features in the Phe-lid. *Acta Crystallogr D Biol Crystallogr* **66**, 953–961.
55 Vertessy BG, Persson R, Rosengren AM, Zeppezauer M & Nyman PO (1996) Specific derivatization of the active site tyrosine in dUTPase perturbs ligand binding to the active site. *Biochem Biophys Res Commun* **219**, 294–300.
56 Saito K, Nagashima H, Noguchi K, Yoshisue K, Yokogawa T, Matsushima E, Tahara T & Takagi S (2014) First-in-human, phase I dose-escalation study of single and multiple doses of a first-in-class enhancer of fluoropyrimidines, a dUTPase inhibitor (TAS-114) in healthy male volunteers. *Cancer Chemother Pharmacol* **73**, 577–583.
57 Miyakoshi H, Miyahara S, Yokogawa T, Endoh K, Muto T, Yano W, Wakasa T, Ueno H, Chong KT, Taguchi J *et al.* (2012) 1,2,3-Triazole-containing uracil derivatives with excellent pharmacokinetics as a novel class of potent human deoxyuridine triphosphatase inhibitors. *J Med Chem* **55**, 6427–6437.
58 Miyahara S, Miyakoshi H, Yokogawa T, Chong KT, Taguchi J, Muto T, Endoh K, Yano W, Wakasa T, Ueno H *et al.* (2012) Discovery of highly potent human deoxyuridine triphosphatase inhibitors based on the conformation restriction strategy. *J Med Chem* **55**, 5483–5496.
59 Pecsi I, Hirmondo R, Brown AC, Lopata A, Parish T, Vertessy BG & Toth J (2012) The dUTPase enzyme is essential in *Mycobacterium smegmatis*. *PLoS ONE* **7**, e37461.
60 Muha V, Horvath A, Bekesi A, Pukancsik M, Hodoscsek B, Merenyi G, Rona G, Batki J, Kiss I, Jankovics F *et al.* (2012) Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *PLoS Genet* **8**, e1002738.
61 Merenyi G, Kovari J, Toth J, Takacs E, Zagyva I, Erdei A & Vertessy BG (2011) Cellular response to efficient dUTPase RNAi silencing in stable HeLa cell lines perturbs expression levels of genes involved in thymidylate metabolism. *Nucleosides, Nucleotides Nucleic Acids* **30**, 369–390.
62 Varga B, Migliardo F, Takacs E, Vertessy GB & Magazu S (2008) Experimental study on dUTPase-inhibitor candidate and dUTPase/disaccharide mixtures by PCS and ENS. *J Mol Struct* **886**, 1–3, 128–135.
63 Hemsworth GR, Moroz OV, Fogg MJ, Scott B, Bosch-Navarrete C, Gonzalez-Pacanowska D & Wilson KS (2011) The crystal structure of the *Leishmania major* deoxyuridine triphosphate nucleotidohydrolase in complex with nucleotide analogues, dUMP, and deoxyuridine. *J Biol Chem* **286**, 16470–16481.
64 Whittingham JL, Leal I, Nguyen C, Kasinathan G, Bell E, Jones AF, Berry C, Benito A, Turkenburg JP, Dodson EJ *et al.* (2005) dUTPase as a platform for antimalarial drug design: structural basis for the selectivity of a class of nucleoside inhibitors. *Structure* **13**, 329–338.
65 Chan S, Segelke B, Lekin T, Krupka H, Cho US, Kim MY, So M, Kim CY, Naranjo CM, Rogers YC *et al.* (2004) Crystal structure of the *Mycobacterium tuberculosis* dUTPase: insights into the catalytic mechanism. *J Mol Biol* **341**, 503–517.
66 Nemeth-Pongracz V, Barabas O, Fuxreiter M, Simon I, Pichova I, Rumlova M, Zabranska H, Svergun D, Petoukhov M, Harmat V *et al.* (2007) Flexible segments modulate co-folding of dUTPase and nucleocapsid proteins. *Nucleic Acids Res* **35**, 495–505.
67 Barabas O, Pongracz V, Kovari J, Wilmanns M & Vertessy BG (2004) Structural insights into the catalytic mechanism of phosphate ester hydrolysis by dUTPase. *J Biol Chem* **279**, 42907–42915.
68 Mustafi D, Bekesi A, Vertessy BG & Makinen MW (2003) Catalytic and structural role of the metal ion in dUTP pyrophosphatase. *Proc Natl Acad Sci USA* **100**, 5670–5675.
69 Pecsi I, Szabo JE, Adams SD, Simon I, Sellers JR, Vertessy BG & Toth J (2011) Nucleotide pyrophosphatase employs a P-loop-like motif to enhance catalytic power and NDP/NTP

discrimination. *Proc Natl Acad Sci USA* **108**, 14437–14442.
70 Pang B, McFaline JL, Burgis NE, Dong M, Taghizadeh K, Sullivan MR, Elmquist CE, Cunningham RP & Dedon PC (2012) Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. *Proc Natl Acad Sci USA* **109**, 2319–2324.
71 Hwang KY, Chung JH, Kim SH, Han YS & Cho Y (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*. *Nat Struct Biol* **6**, 691–696.
72 Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H & Pavlov YI (1996) HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast *Saccharomyces cerevisiae*. *Yeast* **12**, 17–29.
73 Burgis NE & Cunningham RP (2007) Substrate specificity of RdgB protein, a deoxyribonucleoside triphosphate pyrophosphohydrolase. *J Biol Chem* **282**, 3531–3538.
74 Zheng J, Singh VK & Jia Z (2005) Identification of an ITPase/XTPase in *Escherichia coli* by structural and biochemical analysis. *Structure* **13**, 1511–1520.
75 Tchigvintsev A, Tchigvintsev D, Flick R, Popovic A, Dong A, Xu X, Brown G, Lu W, Wu H, Cui H *et al.* (2013) Biochemical and structural studies of conserved Maf proteins revealed nucleotide pyrophosphatases with a preference for modified nucleotides. *Chem Biol* **20**, 1386–1398.
76 Fujishiro T, Tamura H, Schick M, Kahnt J, Xie X, Ermler U & Shima S (2014) Identification of the HcgB enzyme in [Fe]-hydrogenase-cofactor biosynthesis. *Angew Chem Int Ed Engl* **52**, 12555–12558.
77 Stenmark P, Kursula P, Flodin S, Graslund S, Landry R, Nordlund P & Schuler H (2007) Crystal structure of human inosine triphosphatase. Substrate binding and implication of the inosine triphosphatase deficiency mutation P32T. *J Biol Chem* **282**, 3182–3187.
78 Stenmark P, Moche M, Gurmu D & Nordlund P (2007) The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in *Escherichia coli*: implications for the reductive mechanism. *J Mol Biol* **373**, 48–64.
79 Lokanath NK, Pampa KJ, Takio K & Kunishima N (2008) Structures of dimeric nonstandard nucleotide triphosphate pyrophosphatase from *Pyrococcus horikoshii* OT3: functional significance of interprotomer conformational changes. *J Mol Biol* **375**, 1013–1025.
80 Gall AD, Gall A, Moore AC, Aune MK, Heid S, Mori A & Burgis NE (2013) Analysis of human ITPase nucleobase specificity by site-directed mutagenesis. *Biochimie* **95**, 1711–1721.
81 Awwad K, Desai A, Smith C & Sommerhalter M (2013) Structural and functional characterization of a noncanonical nucleoside triphosphate pyrophosphatase from *Thermotoga maritima*. *Acta Crystallogr D Biol Crystallogr* **69**, 184–193.
82 McLennan AG (2006) The Nudix hydrolase superfamily. *Cell Mol Life Sci* **63**, 123–143.
83 Taddei F, Hayakawa H, Bouton M, Cirinesi A, Matic I, Sekiguchi M & Radman M (1997) Counteraction by MutT protein of transcriptional errors caused by oxidative damage. *Science* **278**, 128–130.
84 Lin J, Abeygunawardana C, Frick DN, Bessman MJ & Mildvan AS (1997) Solution structure of the quaternary MutT-M2+-AMPCPP-M2+ complex and mechanism of its pyrophosphohydrolase action. *Biochemistry* **36**, 1199–1211.
85 Xia Z, Azurmendi HF, Lairson LL, Withers SG, Gabelli SB, Bianchet MA, Amzel LM & Mildvan AS (2005) Mutational, structural, and kinetic evidence for a dissociative mechanism in the GDP-mannose mannosyl hydrolase reaction. *Biochemistry* **44**, 8989–8997.
86 Ito R, Hayakawa H, Sekiguchi M & Ishibashi T (2005) Multiple enzyme activities of Escherichia coli MutT protein for sanitization of DNA and RNA precursor pools. *Biochemistry* **44**, 6670–6674.
87 Fujikawa K, Kamiya H, Yakushiji H, Nakabeppu Y & Kasai H (2001) Human MTH1 protein hydrolyzes the oxidized ribonucleotide, 2-hydroxy-ATP. *Nucleic Acids Res* **29**, 449–454.
88 Camacho A, Hidalgo-Zarco F, Bernier-Villamor V, Ruiz-Perez LM & Gonzalez-Pacanowska D (2000) Properties of Leishmania major dUTP nucleotidohydrolase, a distinct nucleotide-hydrolysing enzyme in kinetoplastids. *Biochem J* **346** (Pt 1), 163–168.
89 Leveles I, Németh V, Szabó JE, Harmat V, Nyíri K, Bendes AA, Papp-Kádár V, Zagyva I, Róna G, Ozohanics O *et al.* (2013) Structure and enzymatic mechanism of a moonlighting dUTPase. *Acta Crystallogr D* **D69**, 2298–2308.
90 Khatemi BE, Chung On CC, Chamakura KR & Kuty Everett GF (2013) Complete genome of *Bacillus megaterium* podophage Pony. *Genome Announc* **1**, e00860–13.
91 Lopez MS, Hodde MK, Chamakura KR & Kuty Everett GF (2014) Complete genome of *Bacillus megaterium* podophage Page. *Genome Announc* **2**, e00332–14.
92 Moroz OV, Murzin AG, Makarova KS, Koonin EV, Wilson KS & Galperin MY (2005) Dimeric dUTPases, HisE, and MazG belong to a new superfamily of all-alpha NTP pyrophosphohydrolases

with potential 'house-cleaning' functions. *J Mol Biol* **347**, 243–255.
93 Bernier-Villamor V, Camacho A, Hidalgo-Zarco F, Perez J, Ruiz-Perez LM & Gonzalez-Pacanowska D (2002) Characterization of deoxyuridine 5′-triphosphate nucleotidohydrolase from Trypanosoma cruzi. *FEBS Lett* **526**, 147–150.
94 Harkiolaki M, Dodson EJ, Bernier-Villamor V, Turkenburg JP, Gonzalez-Pacanowska D & Wilson KS (2004) The crystal structure of *Trypanosoma cruzi* dUTPase reveals a novel dUTP/dUDP binding fold. *Structure* **12**, 41–53.
95 Prasad GS (2001) Glycine rich P-loop motif in deoxyuridine pyrophosphatase. *Curr Protein Pept Sci* **2**, 301–311.
96 Hemsworth GR, Gonzalez-Pacanowska D & Wilson KS (2013) On the catalytic mechanism of dimeric dUTPases. *Biochem J* **456**, 81–88.
97 Barabas O, Nemeth V, Bodor A, Perczel A, Rosta E, Kele Z, Zagyva I, Szabadka Z, Grolmusz VI, Wilmanns M *et al.* (2013) Catalytic mechanism of alpha-phosphate attack in dUTPase is revealed by X-ray crystallographic snapshots of distinct intermediates, ^{31}P-NMR spectroscopy and reaction path modelling. *Nucleic Acids Res* **41**, 10542–10555.
98 Zhang J & Inouye M (2002) MazG, a nucleoside triphosphate pyrophosphohydrolase, interacts with Era, an essential GTPase in *Escherichia coli*. *J Bacteriol* **184**, 5323–5329.
99 Lyu LD, Tang BK, Fan XY, Ma H & Zhao GP (2013) Mycobacterial MazG safeguards genetic stability via housecleaning of 5-OH-dCTP. *PLoS Pathog* **9**, e1003814.
100 Goncalves AM, de Sanctis D & McSweeney SM (2011) Structural and functional insights into DR2231 protein, the MazG-like nucleoside triphosphate pyrophosphohydrolase from *Deinococcus radiodurans*. *J Biol Chem* **286**, 30691–30705.
101 Nonaka M, Tsuchimoto D, Sakumi K & Nakabeppu Y (2009) Mouse RS21-C6 is a mammalian 2′-deoxycytidine 5′-triphosphate pyrophosphohydrolase that prefers 5-iodocytosine. *FEBS J* **276**, 1654–1666.
102 Wu B, Liu Y, Zhao Q, Liao S, Zhang J, Bartlam M, Chen W & Rao Z (2007) Crystal structure of RS21-C6, involved in nucleoside triphosphate pyrophosphohydrolysis. *J Mol Biol* **367**, 1405–1412.
103 Lee S, Kim MH, Kang BS, Kim JS, Kim GH, Kim YG & Kim KJ (2008) Crystal structure of Escherichia coli MazG, the regulator of nutritional stress response. *J Biol Chem* **283**, 15232–15240.
104 Hayakawa H, Hofer A, Thelander L, Kitajima S, Cai Y, Oshiro S, Yakushiji H, Nakabeppu Y, Kuwano M & Sekiguchi M (1999) Metabolic fate of oxidized guanine ribonucleotides in mammalian cells. *Biochemistry* **38**, 3610–3614.
105 Nakamura T, Zhao Y, Yamagata Y, Hua YJ & Yang W (2012) Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487**, 196–201.
106 Galperin MY & Koonin EV (2012) Divergence and convergence in enzyme evolution. *J Biol Chem* **287**, 21–28.
107 Jansson K, Blomberg A, Sunnerhagen P & Alm Rosenblad M. Evolutionary loss of 8-oxo-G repair components among eukaryotes. *Genome Integr* **1**, 12.
108 Requena CE, Perez-Moreno G, Ruiz-Perez LM, Vidal AE & Gonzalez-Pacanowska D (2014) The NTP pyrophosphatase DCTPP1 contributes to the homoeostasis and cleansing of the dNTP pool in human cells. *Biochem J* **459**, 171–180.
109 Simone PD, Pavlov YI & Borgstahl GE (2013) ITPA (inosine triphosphate pyrophosphatase): from surveillance of nucleotide pools to human disease and pharmacogenetics. *Mutat Res* **753**, 131–146.
110 Wilson PM, LaBonte MJ, Lenz HJ, Mack PC & Ladner RD (2012) Inhibition of dUTPase induces synthetic lethality with thymidylate synthase-targeted therapies in non-small cell lung cancer. *Mol Cancer Ther* **11**, 616–628.
111 Sakumi K, Abolhassani N, Behmanesh M, Iyama T, Tsuchimoto D & Nakabeppu Y (2010) ITPA protein, an enzyme that eliminates deaminated purine nucleoside triphosphates in cells. *Mutat Res* **703**, 43–50.
112 Stepchenkova EI, Tarakhovskaya ER, Spitler K, Frahm C, Menezes MR, Simone PD, Kolar C, Marky LA, Borgstahl GE & Pavlov YI (2009) Functional study of the P32T ITPA variant associated with drug sensitivity in humans. *J Mol Biol* **392**, 602–613.
113 Gad H, Koolmeister T, Jemth AS, Eshtad S, Jacques SA, Strom CE, Svensson LM, Schultz N, Lundback T, Einarsdottir BO *et al.* (2014) MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **508**, 215–221.
114 Huber KV, Salah E, Radic B, Gridling M, Elkins JM, Stukalov A, Jemth AS, Gokturk C, Sanjiv K, Stromberg K *et al.* (2014) Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **508**, 222–227.
115 Sanchis-Gomar F, Pareja-Galeano H & Lucia A (2014) The hMTH1 paradox: antioxidants recommended in cancer? *DNA Repair (Amst)*. **21**, 163–164.
116 Minasov G, Teplova M, Stewart GC, Koonin EV, Anderson WF & Egli M (2000) Functional implications from crystal structures of the conserved *Bacillus subtilis* protein Maf with and without dUTP. *Proc Natl Acad Sci USA* **97**, 6328–6333.

117 Nakamura T, Meshitsuka S, Kitagawa S, Abe N, Yamada J, Ishino T, Nakano H, Tsuzuki T, Doi T, Kobayashi Y *et al*. (2010) Structural and dynamic features of the MutT protein in the recognition of nucleotides with the mutagenic 8-oxoguanine base. *J Biol Chem* **285**, 444–452.

118 Svensson LM, Jemth AS, Desroses M, Loseva O, Helleday T, Hogbom M & Stenmark P (2011) Crystal structure of human MTH1 and the 8-oxo-dGMP product complex. *FEBS Lett* **585**, 2617–2621.

119 Waterhouse AM, Procter JB, Martin DM, Clamp M & Barton GJ (2009) Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.