Evolution of the TPP-dependent enzyme family

Seán J. Costelloe[1], John M. Ward[2], Paul A. Dalby[1,*]

1. Department of Biochemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK
2. Department of Biochemistry and Molecular Biology, University College London, Gower Street, London

* Corresponding author

p.dalby@ucl.ac.uk

Tel: +44(0)20 7679 2962

Fax: +44(0)20 7383 2348

Footnote title: Evolution of TPP-dependent enzymes

Abstract

The evolutionary relationships of the thiamine pyrophosphate (TPP) dependent family of enzymes was investigated by generation of a neighbor joining (NJ) phylogenetic tree using sequences from the conserved pyrophosphate (PP) and pyrimidine (Pyr) binding domains of 17 TPP-dependent enzymes. This represents the most comprehensive analysis of TPP-dependent enzyme evolution to date. The phylogeny was shown to be robust by comparison with maximum likelihood (ML) trees generated for each individual enzyme. The study discerned the order in which enzymes diverged in the progenote population and several instances where enzyme types diverged later, and broadly confirms the evolutionary history proposed recently from structural comparisons alone (Duggleby 2006). The relationship between the PP- and Pyr- domains and the recruitment of additional protein domains was examined using the transketolase C-terminal (TKC) domain as an example. This domain has been recruited by several members of the family and yet forms no part of the active site and has unknown function. Removal of the TKC-domain was found to increase activity towards β-hydroxypyruvate (HPA) and glycolaldehyde (GA). Further truncations of the Pyr-domain yielded several variants with retained activity. This suggests the influence of TKC-domain recruitment on the evolution of the mechanism and specificity of transketolase (TK) has been minor, and that the smallest functioning unit of TK comprises the PP- and Pyr-domains whose evolutionary histories extend to all TPP-dependent enzymes.

Keywords

Thiamine pyrophosphate, transketolase, domains, phylogeny, mutagenesis, truncation

Introduction

TPP cofactor dependent metabolic enzymes perform a diverse range of reactions, including nonoxidative decarboxylation of α-keto acids, oxidative decarboxylation of α-keto acids, carboligation, as well as cleavage of C-C bonds (Sprenger and Pohl 1999). Their diverse chemistries appear to have evolved by domain rearrangement, gene duplication, gene-splitting, the recruitment of additional domains, and the incremental mutation of functionally important enzyme residues, though the relative contributions of these processes are unknown. The common feature of all TPP-dependent enzymes is the binding of TPP at the interface of a conserved PP- and Pyr- domain. The three-dimensional structures of TK, pyruvate oxidase (PO), and pyruvate decarboxylase (PDC) confirm that their TPP-binding sites are very similar, despite considerable differences in overall primary structure and domain order (Muller et al. 1993). While most TPP-dependent enzymes form active sites between PP- and Pyr- domains on different subunits, pyruvate ferredoxin reductase (PFRD), and more recently D-xylulose-5-phosphate synthase (DXPS), form the active site between PP- and Pyr- domains on the same subunit (Chabriere et al. 1999; Xiang et al. 2007).

Phylogenetic studies provide information about the mechanisms underlying protein change and, where enzymes are similar, their relationship can be used to infer function. Here the TPP-dependent enzymes are classified into six groups (Figure 1) in which the domains are alternatively arranged in the primary structure, or appear on different subunits. Enzymes of the TK-like group (TK; DXPS; dihydroxyacetone synthase (DHAS); and phosphoketolase (PKL)) contain the PP-, Pyr- and TKC- domains on the same subunit, and are homodimeric, except PKL which is a homohexamer (Meile et al. 2001). The 2-oxoisovalerate dehydrogenase (2OXO) -like enzymes form the E1 subunits of large multienzyme complexes, and differ from the TK-like enzymes in that the PP- domain is found on a different subunit to the Pyr- and TKC- domains. PFRD enzymes contain the PP-, Pyr-, TKC-, plus up to four additional

domains, designated D3, D4, D5 and D7, in various arrangements. The PDC-like group which includes: PDC; indolepyruvate decarboxylase (IPDC); phenylpruvate decarboxylase (PhPDC); PO; acetolactate synthase (ALS); glyoxylate carboligase (GXC); benzaldehyde lyase (BAL); oxalyl CoA decarboxylase (OCADC); and benzoylformate decarboxylase (BFDC), are homotetrameric, and contain a transhydrogenase dIII- (TH3) domain between the Pyr- and PP-domains. In addition to TPP, PO, ALS and GXC each require FAD (Bertagnolli and Hager 1993; Cromartie and Walsh 1976; Chang et al. 1993; Koland JG 1982). Finally, SPDC and PPDC both consist only of PP- and Pyr-domains (Graupner et al. 2000). In SPDC, the PP- and Pyr-domains form the separate subunits (α- and β-) of an $\alpha_2\beta_2$-heterotetramer. In PPDC, the PP- and Pyr- domains are fused and form a homotrimer (Zhang et al. 2003).

The PP- and Pyr- domains are thought to have diverged from a common ancestor and then fused into a single chain (Todd et al. 2001). Sulfopyruvate decarboxylase (SPDC) most closely resembles the un-fused ancestor of all TPP-dependent enzymes, while phosphopyruvate decarboxylase (PPDC) is structurally closest to the fused common ancestor, as inferred in recent structural studies (Duggleby 2006). However, further analysis is required at the protein sequence level as variations in the domain recruitment and arrangement, and also the occurrence of either intra- or inter-chain active site formation between PP- and Pyr-domains, may have occurred more than once during the evolutionary history. A previous sequence level evolutionary analysis for the TPP-dependent enzymes (Talarico et al. 2001) examined only five enzyme types with the apparent inclusion of group specific domains which may have introduced considerable bias.

Here an extensive phylogenetic analysis of seventeen TPP-dependent enzyme types (TK; DXPS; DHAS; PKL; 2OXO; PFRD; PDC; IPDC; PhPDC; PO; ALS; GXC; BFDC; BAL; OCADC; SPDC; and PPDC) was performed, based solely on common regions of PP- and Pyr-domain protein sequence to remove the influence of additional sequence and domains. To

account for alternative domain arrangements, the sequences of the PP- and Pyr- domains were aligned independently before fusion into a single PP-Pyr domain alignment and construction of a phylogenetic tree for all TPP-dependent enzymes. Our analysis shows that sequences from only the PP- and Pyr- domain are sufficient to infer the evolutionary history of TPP-dependent enzymes, *albeit* with some guidance from available structures. To assess the impact of additional domain recruitment upon PP- and Pyr- domain evolution, we have examined the impact upon TK activity of removing the highly conserved TKC-domain. This domain is common to the TK-like, 2OXO-like, and PFRD enzymes, and has an undefined function, although a role in TK regulation has been hypothesised (Nikkola et al. 1994). We have also made four further truncations into the Pyr-domain of the TK gene to provide further insight into the relative evolutionary roles of the PP- and Pyr- domains.

Materials and Methods

Choice of enzymes for analysis.

TPP-dependent enzymes were included for study where sufficient sequences with high homology were available to provide a meaningful sequence alignment and clear phylogeny (Green 1989). PFRD domain architectures can vary between species. There are also significant differences in the secondary structures of the catalytic domains between species. For these reasons, and because the crystal structure has been solved, the *Daf*-like PFRDs were chosen for study.

Sequence alignments for the PP- and Pyr- domains

Sequences were retrieved using a *BLAST* (Ye et al. 2006) search with default parameters. Putative sequences returned from BLAST searches were omitted, and the remaining sequences aligned using *ClustalW* (Thompson et al. 1994; Felsenstein 1993).

Alignments for the PP- and Pyr- domains of all the TPP-dependent enzyme groups were generated separately. Pyr-domains were defined as residues 25-111 in *S. cerevisiae* PDC (*Sce*PDC), and residues 323-538 in *E. coli* TK (*Eco*TK), whereas the PP- domains were defined as residues 302-530 in *Sce*PDC, and residues 1-350 in *Eco*TK. The TK-like, PFRDs and 2OXO group alignments were aligned to the *Eco*TK domain boundaries, while the PDC-like and the PPDC/SPDC group alignments were aligned to the *Sce*PDC domain boundaries. The crystal structures of *Eco*TK (1QGD.pdb), *P. putida* 2OXO (*Ppu*2OXO) (2BP7.pdb), *Daf*PFRD (1BOP.pdb), *Sce*PDC (1PVD.pdb) and *L. plantarum* PO (*Lpl*PO) (1POX.pdb), were used to refine the alignment of functionally important residues (Table 1) and secondary structure elements. Alignments were de-gapped to include only residues found in regions of secondary structure common to all TPP-enzymes, ensuring that only the most informative stretches of sequences were used in the phylogenetic analysis. The PP- and Pyr-domain alignments were concatenated prior to phylogenetic analysis.

Phylogenetic analysis of all TPP-dependent enzymes

The concatenated PP- and Pyr- domain alignment was input into the PHYLIP package (Felsenstein 1989) program *Protdist*, to generate a distance matrix which was then input into the program *Neighbor* to construct a NJ tree. Trees were was plotted using *Mega3* (Kumar et al. 2001) and *Treeview* (Page 1996).

Phylogenetic analyses of individual TPP-dependent enzymes

ML phylogenies were generated for each enzyme type individually, allowing comparison with clades in the overall TPP-dependent enzyme tree. Enzymes with PP- and Pyr- domains on separate subunits (SPDC and 2OXO), were concatenated prior to their individual alignments. ML tree phylogenies for each individual enzyme were obtained using the PHYLIP

package program *ProML* (Felsenstein 1993). No phylogenies were constructed for BAL, PhPDC or DHAS since they were represented by only three sequences each.

Conservation and mutagenesis of the TK C-terminal domain

To examine the conservation of the TKC-domain in TK, the sequences for *Eco*TK (1QGD.pdb) and *Sce*TK (1NGS.pdb) were aligned in *Bioedit*. Throughout this paper conservation was defined using the Blosum62 matrix (Henikoff and Henikoff 1992). The TKC-domain is defined as beginning at residue 540 in *Eco*TK (Lindqvist et al. 1992), and so a stop codon was introduced at Gly540 by site-directed mutagenesis (SDM). Further truncation was obtained by introducing stop codons into loop regions of the Pyr-domain, minimising the likelihood of structural disruptions. Truncated TK mutants produced were G540Stop, Glu527Stop, Arg492Stop, His461Stop and Gln453Stop. Site directed mutagenesis was carried out using the Quickchange^{TM} kit (Stratagene), the pQR711 plasmid, which expresses transketolase from the *tktA* gene using its own promoter (French and Ward 1995), and the DNA primers (Qiagen Ltd.), which were as follows (mutagenic codons in bold):

G540stop: CGCGCGCGGT**TGA**TATGTGCTG;

E527stop: GGCGCAGCAG**TAA**CGAACTGAAG;

R492stop: GTCTACATGG**TGA**CCGTGTGAC;

H461stop: GGTTTACACC**TAG**GACTCCATCGG;

Q453stop: GCTGATGAAA**TAG**CGTCAGGTGATG.

Mutant plasmids were transformed into XL10-Gold cells (Stratagene) and confirmed by DNA sequencing. Individual colonies were selected and grown in 10 ml Luria-Bertani (LB) with 150 μg ml^{-1} ampicillin for 16 hrs at 200 rpm, 37 °C. 8 ml of culture was centrifuged at 4,000 g for 10 mins and resuspended in 2 ml of 5 mM sodium phosphate buffer, pH 7.0, before

sonication for 10 cycles of 10 seconds. The lysate was clarified by centrifuging at 4000 g for 10 mins.

Enzyme activity of the wild-type and truncated TK mutants

The TK catalysed reaction of HPA and GA was performed in a total volume of 600 µl, with 50 mM Tris-HCl, pH 7.0, 50 mM HPA, 50 mM GA, 9 mM $MgCl_2$, 2.4 mM TPP, and 50% clarified lysate solution at a final enzyme concentration of 0.5 mg ml^{-1} determined with an Agilent 2100 bioanalyser. Reagents were mixed and incubated for 30 mins at room temperature with the cofactors prior to initiation by addition of the GA. The reaction at 25 ºC was monitored by removing 100 µl samples at 5, 15, 30, 60 and 120 minutes and quenching with 100 µl of 0.2 % trifluoroacetic acid (TFA). The concentration of L-erythrulose product was determined by HPLC (Dionex Corp.) of 10 µl samples injected onto a 300 mm Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories) maintained at 60 ºC, a mobile phase of 0.1 % (v/v) TFA, and a flow rate of 0.6 ml min^{-1}. Product was quantified by electrochemical detection (ECD).

Results and Discussion

Alignment and conservation within PP- and Pyr- domains

The case for divergent evolution is strong for sequences with over 30 % identity, particularly where strong similarities also exist between three-dimensional structures and biological function (Chung and Subbiah 1996). By contrast, for less than 30 % identity, inference of evolutionary relationships between enzymes is difficult with sequence data alone and requires a comparison of structures and functionally important residues to guide the alignment. The sequence homology between different TPP-dependent enzyme groups is typically low, for example the PP- and Pyr- domain sequences of *Eco*TK and *Sce*PDC share

~15 % homology. Within each enzyme group, sequence homology is high enough to directly infer evolutionary relationships. Therefore, we have aligned enzymes using sequence homology within groups of the same enzyme architecture, then refined and combined them by structural comparison of enzymes from different groups (Figures 2 & 3). In this way, we have elucidated regions of secondary structure that are common to all TPP-dependent enzymes and inferred phylogeny based on the sequence homology within these regions.

Structural similarity between the different TPP-dependent enzyme groups is observed in the PP- and Pyr- domains. A comparison of the PP- domains of the five representative structures (Figure 2) shows that six α-helices and five β-strands are common to these enzymes. Similarly, comparison of the Pyr-domains shows that five α-helices and four β-strands are common (Figure 3). Residues known to make important contributions to the activity of various members of the TPP-dependent enzymes are also shown in Table 1 along with the degree of conservation observed for each enzyme type in our sequence alignment. Their good alignment inspires confidence in the subsequent phylogenetic analysis and strongly agrees with divergence of the enzyme groups from a distant common ancestor. The alignment also provides a useful summary of the level of conservation of functionally important residues across the members of the enzyme family.

Residue Ile189 (*Eco*TK numbering unless stated otherwise) is directly above the face of the thiazolium ring of TPP and interacts with the divalent metal ion, which in turn anchors the diphosphate group of TPP (Meshalkina et al. 1997). This residue is highly conserved across TK, DXPS and PKL, but less conserved in the other enzyme types. Glu477 (*Sce*PDC) is similarly involved in metal-ion binding and is essential for catalysis in PDC (Fiedler et al. 2002). This residue is highly conserved in PDC and IPDC but only occurs in 33% of PhPDCs.

Residues Thr991, Val993 and Tyr994 (*Daf*PFRD) are known to coordinate the metal ion in PFRD (Chabriere et al. 1999). Thr991 is conserved in all PFRDs examined. Val993 is

similarly conserved in PFRD, PPDC, SPDC and PO. Tyr994 is conserved in all PFRDs, as well as occurring in nearly 58% of PO sequences. Residue Tyr244 of *Ppu*2OXO is also involved in metal-binding (Ævarsson et al. 1999), and is found in all 2OXO, but occurs in no other enzymes.

Residues Gly154, Asp155 and Asn185 form part of the "ThDP-binding" motif of TK (Schenk et al. 1997) and are highly conserved across all seventeen enzyme types, reflecting their critical roles in TPP-binding. Phe434, Phe437, and Tyr440 form a hydrophobic pocket surrounding the thiazolium ring of TPP (Kochetov 2001). Phe434, is conserved in TK, DXPS, DHAS, PKL, PDC, IPDC and PhPDC. Phe437 is conserved in TK, DXPS, DHAS and 2OXO, whereas Tyr440 is conserved in TK, DHAS and 2OXO.

Certain residues, with proposed roles in TPP-binding were found to be conserved specifically among members of the PDC-like group of TPP-dependent enzymes. Tyr474 (Hasson et al. 1998) (*Sce*PDC), for example is highly conserved in PDC, SPDC, BFDC, IPDC, PhPDC and BAL. This position is conserved in only 17% of PO sequences, 53% of ALS sequences and 13% of PPDC sequences, suggesting it is not strictly essential in these enzymes.

Asp381 is involved in TPP binding (Meshalkina et al. 1997) and is conserved only in TK and 2OXO, as is His461 which has a role in phosphate binding (Meshalkina et al. 1997). Ser385, also believed to be involved in phosphate binding is conserved in TK (Nilsson et al. 1997), PFRD, and OCADC. Finally, the phosphate binding Arg520 (Nilsson et al. 1997) is conserved in TK, DXPS, DHAS, PKL, 2OXO, PPDC, and PhPDC.

Glu411 is essential for catalysis (Kern et al. 1997), and is 100% conserved in all enzymes. His26 and His66 are both directly involved in catalysis and the substrate specificity of TK (Wikner et al. 1997), and are found to be highly conserved in all TK-like

enzymes. His100 is conserved in all TK-like enzymes except DXPS, agreeing with an earlier report (Fiedler et al. 2002).

Finally, His473 is involved in transition-state stabilisation (Wikner et al. 1997) and is conserved in TK, DXPS, DHAS, PKL, PFRD, 2OXO, PDC, IPDC, PhPDC and PPDC. Glu160 is part of a hydrogen-bonding network in TK (Meshalkina et al. 1997) and is highly conserved in TK, DHAS, 2OXO and PFRD. Asp469 determines the stereospecificity (Nilsson et al. 1997) of TK and is highly conserved in TK, DXPS, DHAS, BAL and OCADC.

Phylogeny of the TPP-dependent enzymes

The phylogeny generated using the NJ method for the 371 TPP-dependent enzymes, is remarkably well resolved as seen in the overview tree in Figure 4. A fully detailed version is available upon request from the authors. Figure 5 shows a highly simplified version of the overall tree, while Figure 6 summarises the clustering of sequences by enzyme and taxonomy, and collectively they show that sequences cluster well with respect to domain arrangement, enzyme type and taxonomy. While an initial phylogenetic analysis of 52 TK sequences using a variety of tree-building methods found that the most computationally intensive ML method produced the phylogeny which best agreed with the universal tree of life (data not shown), the individual enzyme clades in the overall NJ tree generally grouped in the same way as for the individual enzyme ML trees. Thus, the regions of secondary structure common to all TPP-dependent enzyme families produce a powerful phylogenetic signal.

The phylogeny is broadly grouped into three sections with the first formed from a clade containing SPDC and PPDC enzymes within which they have clustered well (Figures 4 & 5). This first clade is consistent with the evolutionary ancestor of all TPP-dependent enzymes as proposed previously based on structural comparison alone (Duggleby 2006). In the second part, the PDC-like enzymes form one large grouping within which PO, OCADC, BAL, and

BFDC enzymes form distinct clades (Figure 6). A distinct GXC clade is also formed, though this is found within the ALS grouping and splitting it into two clades which were also observed in the individual ML tree for ALS (data not shown). This finding suggests that GXC has evolved from an ALS-like enzyme. Furthermore, the GXC sequences have a very high level of homology with an average of 65% sequence identity, compared to the other enzymes which have average sequence identities in the range 30% (BFDC) to 47% (DXPS). This suggests that the enzyme is relatively recently evolved, while the ability of GXC (and PO) to perform the ALS reaction at a low rate (Chang et al. 1993) suggests that such an ancestry is plausible. The GXC reaction is possibly a promiscuous function of ALS, which has become the primary reaction in GXC after gene duplication.

All of the PDC enzymes form a distinct clade, except the *M. tuberculosis* (*Mtu*) PDC which groups with the IPDC from the same organism, and the IPDCs and PhPDCs, which form two mixed clades close to the PDC enzymes with one that is more evolutionarily related to PDC than the other. These observations highlight the care that is needed when classifying enzymes of similar type. It would seem that many PDC and IPDC enzymes have been classified on their ability to catalyse one certain reaction. It has been suggested that the use of indolepyruvate depends on active-site volume, where the bulky indole moiety is accommodated by IPDC, but excluded by three large, hydrophobic amino acids in PDC (Schutz et al. 2003). However, it may be the case that many IPDCs can also decarboxylate pyruvate. It is less clear which of the PDCs, apart from *Sce*PDC (Schutz et al. 2003) have activity towards indolepyruvate. The fact that PhPDC can utilise phenylpyruvate or indolepyruvate in different metabolic pathways suggests that perhaps the promiscuous indolepyruvate usage by PhPDC may, after gene duplication, have led to the emergence of the IPDC enzyme, although this is highly speculative.

The third major part of the overall phylogeny groups together the TK-like, 2OXO and PFRD enzymes, confirming the common ancestry that was suggested by their common TKC-domain. The PFRD clade is distinct, suggesting that it split from the other TKC-containing enzymes before they in turn diverged from a more recent common ancestor. The positioning of 2OXO enzymes within the TK-like enzymes suggests that the enzyme evolved from a TK-like enzyme, sharing its most recent common ancestor in the phylogeny with the PKLs (Figures 5 and 6). This would suggest that the TPP-enzymes which are part of multi-enzyme complexes (eg. 2OXO), evolved from a homodimeric TK-like ancestor which was recruited into the complex. This result differs from that proposed previously using a purely structural comparison, where divergence to form the multi-enzyme complex was proposed to occur before the emergence of the TK-like and PFRD enzymes from a heterotetrameric ancestor enzyme containing the TKC-domain (Duggleby 2006).

Within the PFRD / TK-like / 2OXO grouping, all the DXPS and PKL sequences formed two distinct clades. However, the DHAS sequences did not form a distinct cluster and were distributed within the main TK clade suggesting that the ability to use formaldehyde as a substrate has evolved several times in TK. The TK sequences also grouped together (with the DHAS interspersed), although two outliers from *B. mellitensis biovar abortis (Bme)* and *C. acetobutylicum* (*Cac*2), appeared respectively in the clades for 2OXO and PKL. These two sequences were also found as outliers in the individual TK ML phylogeny (data not shown).

Using data from this study, the evolutionary history of the TPP-dependent enzyme family can be assembled (Figure 7). This history broadly confirms the one described recently by Duggleby, which was based on structural comparisons alone (Duggleby 2006), with the exception of the evolutionary route of the 2OXO type enzymes, as discussed above. The earliest ancestor of the TPP-dependent enzymes may have contained regions for both the pyrophosphate and pyrimidine regions of the TPP molecule. Dimerisation of this enzyme

would have improved catalysis perhaps through the emergence of further stabilising interactions with the TPP cofactor. Gene duplication followed by divergence may then have resulted in specialised PP- and Pyr- domains which each bound different parts of the TPP molecule. The corresponding $\alpha_2\beta_2$-heterotetramer assembly survives in modern SPDC enzymes. Fusion of the PP- and Pyr- domain genes subsequently resulted in the homodimeric architecture (Figure 7), with subunits similar to modern PPDC enzymes, although PPDC has been reported as adopting a homotrimeric state (Zhang et al. 2003). All other TPP-dependent enzymes appear to have evolved from a common homodimeric ancestor.

The TH3-domain is recruited by the ancestor of the PDC-like enzymes, from which the PDC clade evolves. TH3-domain recruitment may also have coincided with adoption of the homotetrameric assembly observed for the PDC clade. Recruitment of the TKC-domain by an ancient enzyme as shown in Figure 7 (Box C) formed the ancestor of the TK-like, 2OXO-like and the PFRD enzymes. It appears that PFRD was the first TK-like enzyme to diverge with the recruitment of at least three (four in the case of the *Daf*-like PFRDs) additional domains. Modern TK-like enzymes then emerge, from which the 2OXO enzymes emerge by splitting the subunit across two genes with the PP- domain on one subunit and the Pyr- and TKC- domains on the other subunit (Figure 1). At the same time as the emergence of 2OXO, the PKL enzymes seem to have evolved, adopting a homohexameric structure.

The crystal structure of DXPS has been recently published (Xiang et al. 2007), showing that DXPS binds TPP between PP- and Pyr- domains on the same chain. Previously, it was believed that PFRD was the only TPP-dependent enzyme to bind cofactor in this intra-chain manner. The recent DXPS structure allows us to speculate that the evolutionary switch between inter- and intra- chain binding may be relatively common and can arise through more than one mechanism. The intra-chain TPP-binding in DXPS is facilitated by the absence of an elongated linker region between the PP- and Pyr- domains, found in other TK-like enzymes (95

residues in TK). By contrast it may be speculated that the intra-chain TPP-binding in PFRD was facilitated by the recruitment of three additional domains between the PP- and Pyr-domains.

Effect of TKC-domain removal

The 140 residue TK C-terminal (TKC) domains of *Eco*TK and *Sce*TK, share 40 identical residues and a further 11 with a similar amino acid (as defined by the Blosum62 matrix). Such conservation suggests that their recruitment was important for the function of transketolase, perhaps for regulation of activity as previously suggested (Nikkola et al. 1994). However, the catalytic mechanism and substrate specificity of TK is determined entirely by the PP- and Pyr-domains so it is possible that the basic TK function may have evolved prior to TKC-domain recruitment. We tested this possibility by removal of the TKC-domain.

Specific activities for the reaction between HPA and GA at pH 7.0, catalysed by *Eco*TK, G540Stop, E527Stop, R492Stop, H461Stop and H453Stop, are shown in Figure 8. Removal of the TKC-domain, without removal of any Pyr-domain or active-site residues, increased the specific activity almost 3-fold, and suggests that this domain may indeed regulate TK activity *in vivo*. The TKC-domain is unlikely to contribute to the stability of the TK molecule, as there are few inter-domain contacts between TKC-domains in the TK dimer (Lindqvist et al. 1992). However, the effect of truncation on stability cannot be ruled out. It is possible that removal of the TKC-domain causes some structural changes in the PP- and Pyr- domains, leading to an increase in the TK reaction rate. There is also a water channel at the interface of the TKC-domains, which is linked to the tunnel connecting the two active sites (Nikkola et al. 1994). Control of this channel, which is disrupted in the truncated mutants, may in some way regulate the enzyme activity.

The retention of function after removal of the TKC-domain highlights the relatively low impact upon the core TPP-dependent activity that recruitment of additional domains can have, and may have significant implications for other TPP-dependent enzymes. Only the PP- and Pyr- domains are required for TK catalysis which is consistent with an evolutionary history in which the most ancient TPP-dependent enzymes consisted only of PP- and Pyr- domains (Figure 7).

Effect of Pyr-domain truncation mutants

Further truncation into the Pyr-domain was also performed to examine the domain boundary. The extent of each truncation in the TK structure is shown for illustrative purposes in Figure 9. Removal of residues 527-539 in E527Stop involves the loss of twelve Pyr-domain residues, eight of which (Tyr541, Thr557, Gly558, Ser559, Glu560, Val561, Ser582 and Glu612) are 100% conserved by sequence similarity or identity. The loss of these residues resulted in a sharp decrease in specific activity relative to G540Stop the mutant, to give 60% of the specific activity of wild-type *Eco*TK. The R492Stop mutation removed six additional highly conserved residues (Arg492, Asp495, Glu498, Ser519, Arg520 and Gln521), and decreased the specific activity to only 7 % that of *Eco*TK. Arg492 and Asp495 form part of the TK motif (Schenk et al. 1997), whereas Arg520 (equivalent to Arg528 in yeast TK), interacts with the phosphate group of natural substrates at the active site entrance (Nilsson et al. 1997). However, the substrates used in this study were not phosphorylated, and so disruption of the TK motif is the most plausible reason for loss of activity.

Specific activity is reduced to just 3% of the wild type in the H461Stop mutation which removes a total of eleven fully conserved residues, including most of the TK motif, and nine residues in a region of the TK active-site that only affects substrate binding. Residue His461 at the active site entrance interacts with the phosphate group of natural substrates (Nilsson et

al. 1997), whereas Asp469 interacts with the C2-hydroxyl group of the ketol donor substrate and is the key determinant of the enantioselectivity of the enzyme (Nilsson et al. 1998), although the mutation D469A does not decrease the specific activity for reaction with the non-phosphorylated substrates used here (Hibbert & Dalby, unpublished). It is likely that the activity retained in the H461Stop mutant is no longer enantioselective. The most probable cause for the loss of specific activity in the H461Stop mutation is the removal of His473 which is conserved in all non-mammalian TK enzymes examined, and has been shown by mutagenesis to significantly affect the K_m of the donor substrate (Wikner et al. 1997).

The increasingly significant loss of conserved residues, phosphate-binding residues, and other substrate binding residues, led to sequentially reduced activity. However, other factors may have led to the observed losses of activity, including decreased enzyme stability, or local structural rearrangements due to the disruption of structural elements.

The H453Stop mutation removed no further conserved or substrate-binding residues, and results in a total deletion of 87 of the 220 residue Pyr-domain, and a remarkable increase in specific activity to 160% of the wild type. This reversion of activity relative to the H461Stop mutation indicates the restored affinity for substrate, speculatively by increasing the accessibility of substrate to the active site, and hence also the association rate constant.

These observations have interesting evolutionary implications. Firstly, up to 40 % of the Pyr-domain can be removed while retaining TK-like function, though most likely without specificity for phosphorylated substrates, or stereoselectivity. It could thus be hypothesised that the PP-domain is the more essential domain for the core catalytic activity. More speculatively, the single-domain TPP-dependent enzyme of the progenitor (Figure 7, Box A) may have resembled the modern PP-domain more than the modern Pyr-domain. After duplication and divergence for transketolase-like activity, the Pyr-domain may then have evolved for improved enantioselectivity and substrate specificity, in particular for

phosphorylated substrates, while recruitment of the TKC-domain may have provided a regulatory mechanism for the enzyme. However, it is important to note that all of the highly conserved Pyr-domain residues which interact with the pyrimidine ring of TPP, have been retained in the truncation mutants.

Conclusions

Within the sequences of a given TPP-dependent enzyme, sequence homology is sufficiently high to confidently infer phylogeny. In compiling the PP- and Pyr- domain alignments, regions of equivalent secondary structure, common to all of the TPP-dependent enzymes examined, were identified and shown to contain strong phylogenetic signals even after such a long period of evolutionary time. In all but a few instances (GXC, DHAS, PhPDC) the enzyme types form into distinct clades, thus their common ancestor is likely to have existed in the progenote population.

Our phylogeny agrees broadly with one deduced using only structural comparisons (Duggleby 2006), except for the 2OXO enzymes which were proposed previously to have been recruited to a multi-enzyme complex prior to the divergence of the TK-like and PFRD enzymes from an ancestor enzyme containing the TKC-domain. Our results suggest that 2OXO evolved later from a TK-like ancestor which was then recruited into the complex.

The phylogeny obtained from only the PP- and Pyr- domains groups together all of the TKC-domain containing enzymes and suggests a common ancestor. Using sequence data alone it is impossible to deduce whether TK-like activity evolved first or whether this evolved in an ancestor that recruited the TKC-domain first. However, improvement of the specific activity for an enzyme mutant containing only the PP- and Pyr- domains suggests that the basic TK mechanism could have evolved prior to TKC-domain recruitment. While the *in vivo* role of the TKC-domain remains unresolved, a regulatory role remains a possibility. In addition to the

TKC-domain, 40% of the Pyr-domain, including the TK motif and 10 active-site residues can be removed and activity retained at a level 4-fold higher than activity in *Eco*TK.

Reference List

Ævarsson, A., Seger, K., Turley, S., Sokatch, J. R., and Hol, W. G. (1999) Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. Nat Struct Biol 6:785-792

Bertagnolli, B. L. and Hager, L. P. (1993) Role of Flavin in Acetoin Production by 2 Bacterial Pyruvate Oxidases. Arch Biochem Biophys 300:364-371

Chabriere, E., Charon, M. H., Volbeda, A., Pieulle, L., Hatchikian, E. C., and Fontecilla-Camps, J. C. (1999) Crystal structures of the key anaerobic enzyme pyruvate : ferredoxin oxidoreductase, free and in complex with pyruvate. Nature Struct Biol 6:182-190

Chang, Y. Y., Wang, A. Y., and Cronan, J. E. (1993) Molecular-Cloning, Dna Sequencing, and Biochemical Analyses of Escherichia-Coli Glyoxylate Carboligase - An Enzyme of the Acetohydroxy Acid Synthase-Pyruvate Oxidase Family. J Biol Chem 268:3911-3919

Chung, S. Y. and Subbiah, S. (1996) A structural explanation for the twilight zone of protein sequence homology. Structure 4:1123-1127

Cromartie, T. H. and Walsh, C. T. (1976) Escherichia coli glyoxalate carboligase. Properties and reconstitution with 5-deazaFAD and 1,5-dihydrodeazaFADH2. J Biol Chem 251:329-333

Duggleby, R. G. (2006) Domain relationships in thiamine diphosphate-dependent enzymes. Acc Chem Res 39:550-557

Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166

Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c.

Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., Konig, S., and Schneider, G. (2002) Snapshot of a key intermediate in enzymatic thiamin catalysis: crystal structure of the alpha-carbanion of (alpha,beta-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*. P Natl Acad Sci USA 99:591-595

French, C. and Ward, J. M. (1995) Improved production and stability of *E. coli* recombinants expressing transketolase for large scale biotransformation. Biotechnol Lett 17:247-252

Graupner, Marion, Xu, Huimin, and White, Robert H. (2000) Identification of the Gene Encoding Sulfopyruvate Decarboxylase, an Enzyme Involved in Biosynthesis of Coenzyme M. J Bacteriol 182:4862-4867

Green, Jeremy B. A. (1989) Pyruvate decarboxylase is like acetolactate synthase (ILV2) and not like the pyruvate dehydrogenase E1 subunit. FEBS Letters 246:1-5

Hasson, M. S., Muscate, A., McLeish, M. J., Polovnikova, L. S., Gerlt, J. A., Kenyon, G. L., Petsko, G. A., and Ringe, D. (1998) The crystal structure of benzoylformate decarboxylase at 1.6 angstrom resolution: Diversity of catalytic residues in thiamin diphosphate-dependent enzymes. Biochemistry 37:9918-9930

Henikoff, S. and Henikoff, J. G. (1992) Amino-Acid Substitution Matrices from Protein Blocks. P Natl Acad Sci USA 89:10915-10919

Kern, D., Kern, G., Neef, H., Tittmann, K., Killenberg-Jabs, M., Wikner, C., Schneider, G., and Hubner, G. (1997) How thiamine diphosphate is activated in enzymes. Science 275:67-70

Kochetov, G. A. (2001) Functional flexibility of the transketolase molecule. Biochemistry-Moscow 66:1077-1085

Koland JG, Gennis RB. (1982) Proximity of reactive cysteine residue and flavin in Escherichia coli pyruvate oxidase as estimated by fluorescence energy transfer. Biochemistry 21:4438-4442

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245

Lindqvist, Y., Schneider, G., Ermler, U., and Sundstrom, M. (1992) 3-Dimensional Structure of Transketolase, A Thiamine Diphosphate Dependent Enzyme, at 2.5 Angstrom Resolution. Embo J 11:2373-2379

Meile, Leo, Rohr, Lukas M., Geissmann, Thomas A., Herensperger, Monique, and Teuber, Michael (2001) Characterization of the D-Xylulose 5-Phosphate/D-Fructose 6-Phosphate Phosphoketolase Gene (xfp) from Bifidobacterium lactis. J Bacteriol 183:2929-2936

Meshalkina, L., Nilsson, U., Wikner, C., Kostikowa, T., and Schneider, G. (1997) Examination of the thiamin diphosphate binding site in yeast transketolase by site-directed mutagenesis. Eur J Biochem 244:646-652

Muller, Y. A., Lindqvist, Y., Furey, W., Schulz, G. E., Jordan, F., and Schneider, G. (1993) A Thiamin Diphosphate Binding Fold Revealed by Comparison of the Crystal-Structures of Transketolase, Pyruvate Oxidase and Pyruvate Decarboxylase. Structure 1:95-103

Nikkola, M., Lindqvist, Y., and Schneider, G. (1994) Refined structure of transketolase from *Saccharomyces cerevisiae* at 2.0 A resolution. J Mol Biol 238:387-404

Nilsson, U., Hecquet, L., Gefflaut, T., Guerard, C., and Schneider, G. (1998) Asp477 is a determinant of the enantioselectivity in yeast transketolase. FEBS Lett 424:49-52

Nilsson, U., Meshalkina, L., Lindqvist, Y., and Schneider, G. (1997) Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. J Biol Chem 272:1864-1869

Page, R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357-358

Schenk, G., Layfield, R., Candy, J. M., Duggleby, R. G., and Nixon, P. F. (1997) Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase. J Mol Evol 44:552-572

Schutz, A., Sandalova, T., Ricagno, S., Hubner, G., Konig, S., and Schneider, G. (2003) Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from Enterobacter cloacae, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. Eur J Biochem 270:2312-2321

Sprenger, G. A. and Pohl, M. (1999) Synthetic potential of thiamin diphosphate-dependent enzymes. J Molec Catal B: Enzym 6:145-159

Talarico, L. A., Ingram, L. O., and Maupin-Furlow, J. A. (2001) Production of the Gram-positive Sarcina ventriculi pyruvate decarboxylase in Escherichia coli. Microbiology 147:2425-2435

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

Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) Evolution of function in protein superfamilies, from a structural perspective. J Mol Biol 307:1113-1143

Wikner, C., Nilsson, U., Meshalkina, L., Udekwu, C., Lindqvist, Y., and Schneider, G. (1997) Identification of catalytically important residues in yeast transketolase. Biochemistry 36:15643-15649

Xiang, S., Usunow, G., Lange, G., Busch, M., and Tong, L. (2007) Crystal structure of 1-deoxy-d-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. J Biol Chem 282:2676-2682

Ye, Jian, McGinnis, Scott, and Madden, Thomas L. (2006) BLAST: improvements for better sequence analysis. Nucl Acids Res 34:W6-W9

Zhang, G. F., Dai, J. Y., Lu, Z. B., and Dunaway-Mariano, D. (2003) The phosphonopyruvate decarboxylase from Bacteroides fragilis. J Biol Chem 278:41302-41308

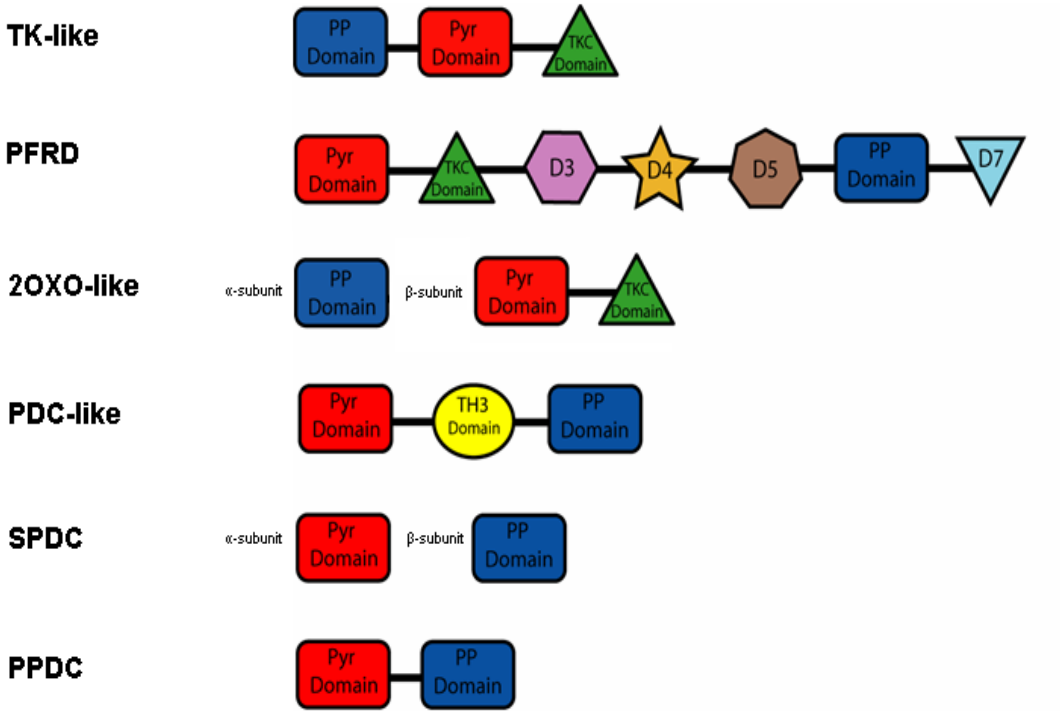

Figure 1: Domain arrangements found in the TPP-dependent enzyme family. Members of the TPP-dependent enzyme family can be classified by their domain architecture. All members contain the catalytic PP- and Pyr- domains which bind TPP. α and β subunits denote cases where the PP- and Pyr- domains are found on different subunits. TPP is bound at the interface of two subunits, except in PFRD and DXPS where TPP is bound by PP- and Pyr- domains on the same chain.

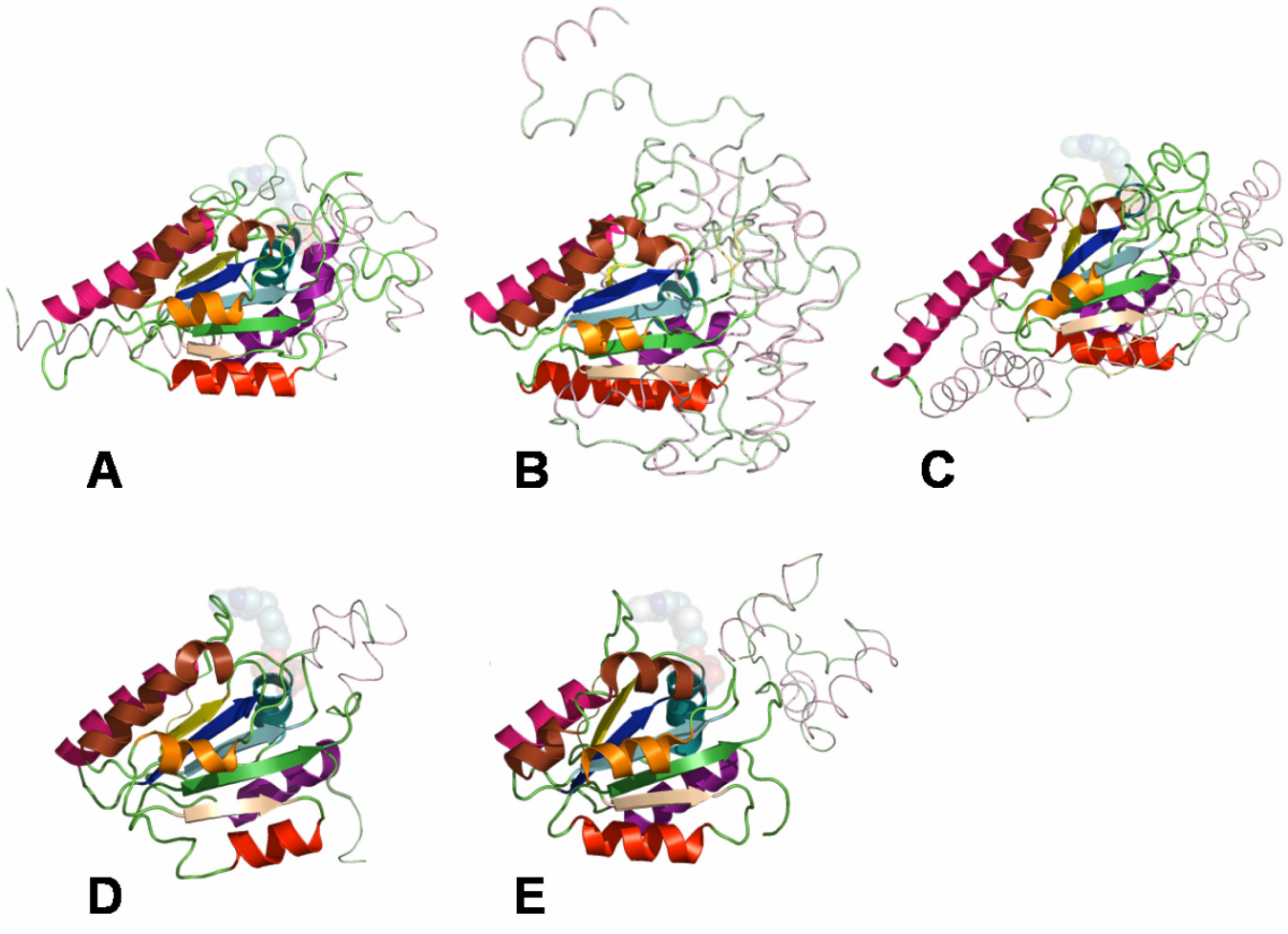

Figure 2: Structural equivalence of PP- domains in TPP-dependent enzymes. Six equivalent α-helices and five equivalent β-strands are each shown in the same colour for: A) TK (1QGD.pdb), B) 2OXO (2BP7.pdb), C) PFRD (1BOP.pdb), D) PDC (1PVD.pdb) and E) PO (POX.pdb). Images were generated in *Pymol* (DeLano, W.L. (2002), The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org).

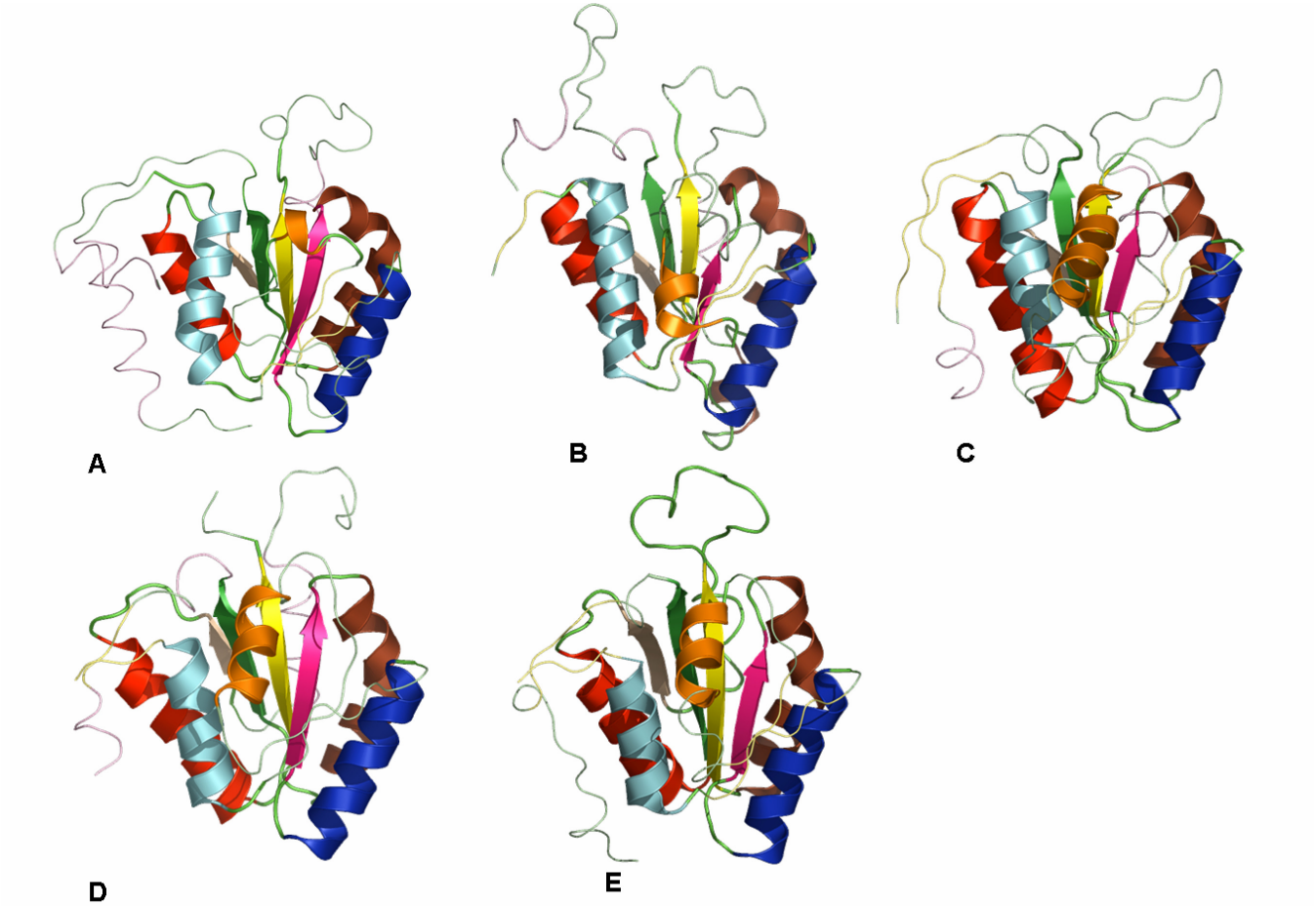

Figure 3: Structural equivalence of Pyr-domains in TPP-dependent enzymes. Five equivalent α-helices and four equivalent β-strands are each shown in the same colour for A) TK (1QGD.pdb), B) 2OXO (2BP7.pdb), C) PFRD (1BOP.pdb), D) PDC (1PVD.pdb) and E) PO (POX.pdb). Images were generated in *Pymol*.

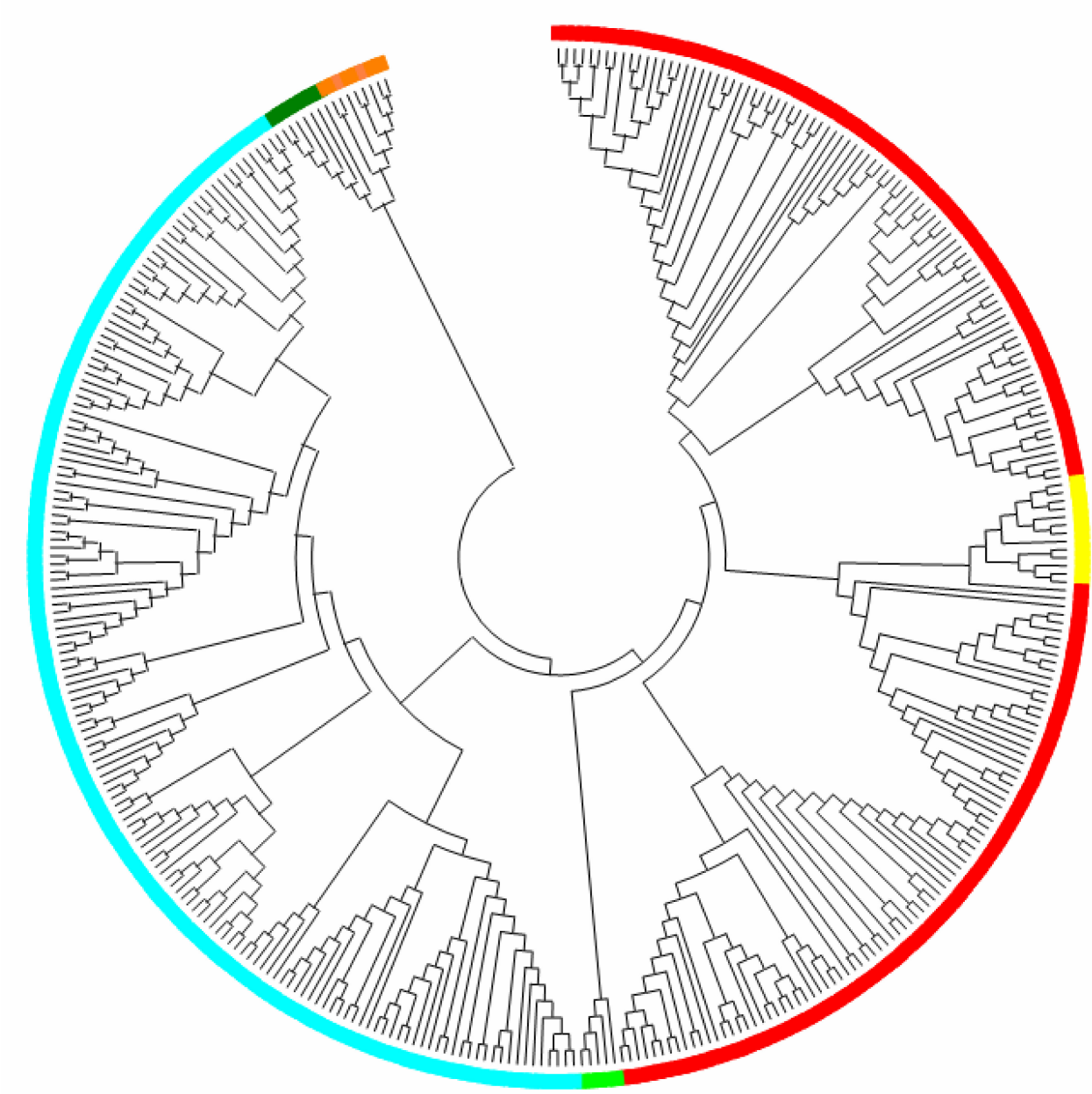

Figure 4: Overview of the TPP-dependent enzyme phylogeny coloured by enzyme group. Enzymes are coloured according to the group of TPP-dependent enzymes to which they belong, as defined in Figure 1. TK-like enzymes are coloured red, PDC-like enzymes are in cyan, PFRD enzymes in the phylogeny are coloured light green, the 2OXO enzymes are in yellow, SPDC enzymes are dark green, while the PPDC enzymes are in orange.

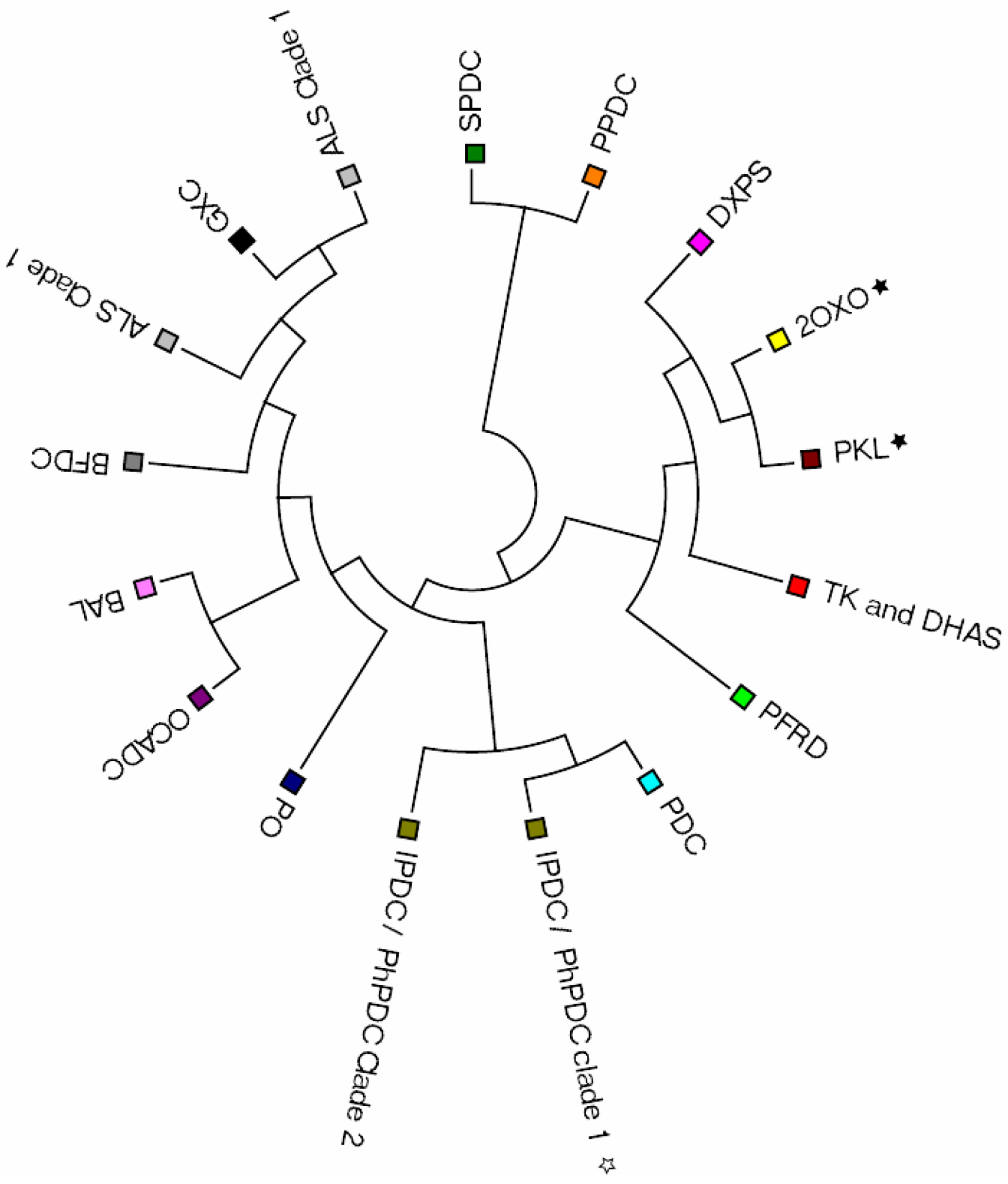

Figure 5: Simplified version of the NJ tree for the TPP-dependent enzymes. Stars indicate where outliers are found. The black stars indicate where *Cac*2TK and *Bme*TK cluster respectively, with the PKL and 2OXO enzymes. The white star indicates where *Mtu*PDC groups within one of the IPDC / PhPDC clades.

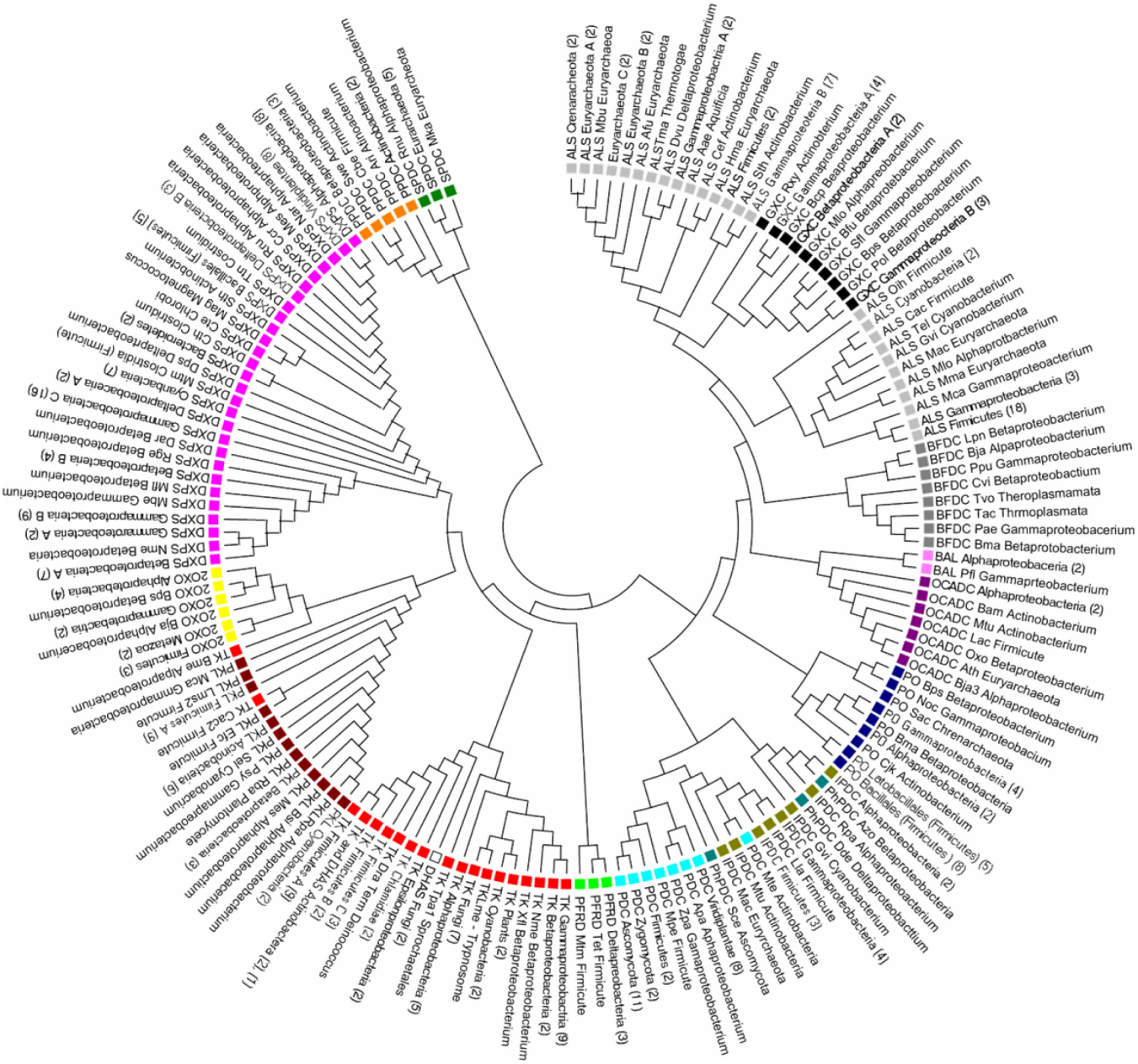

Figure 6: Overall NJ tree for the TPP-dependent enzymes simplified according to enzyme type and taxonomy. Enzyme types are coloured as in Figure 5. For a given enzyme, where more than one species of the same taxon were found to cluster, the clade is collapsed and details given of the class and number of species in that clade. A complete list of the three letter species abbreviations used is available on request from the authors.

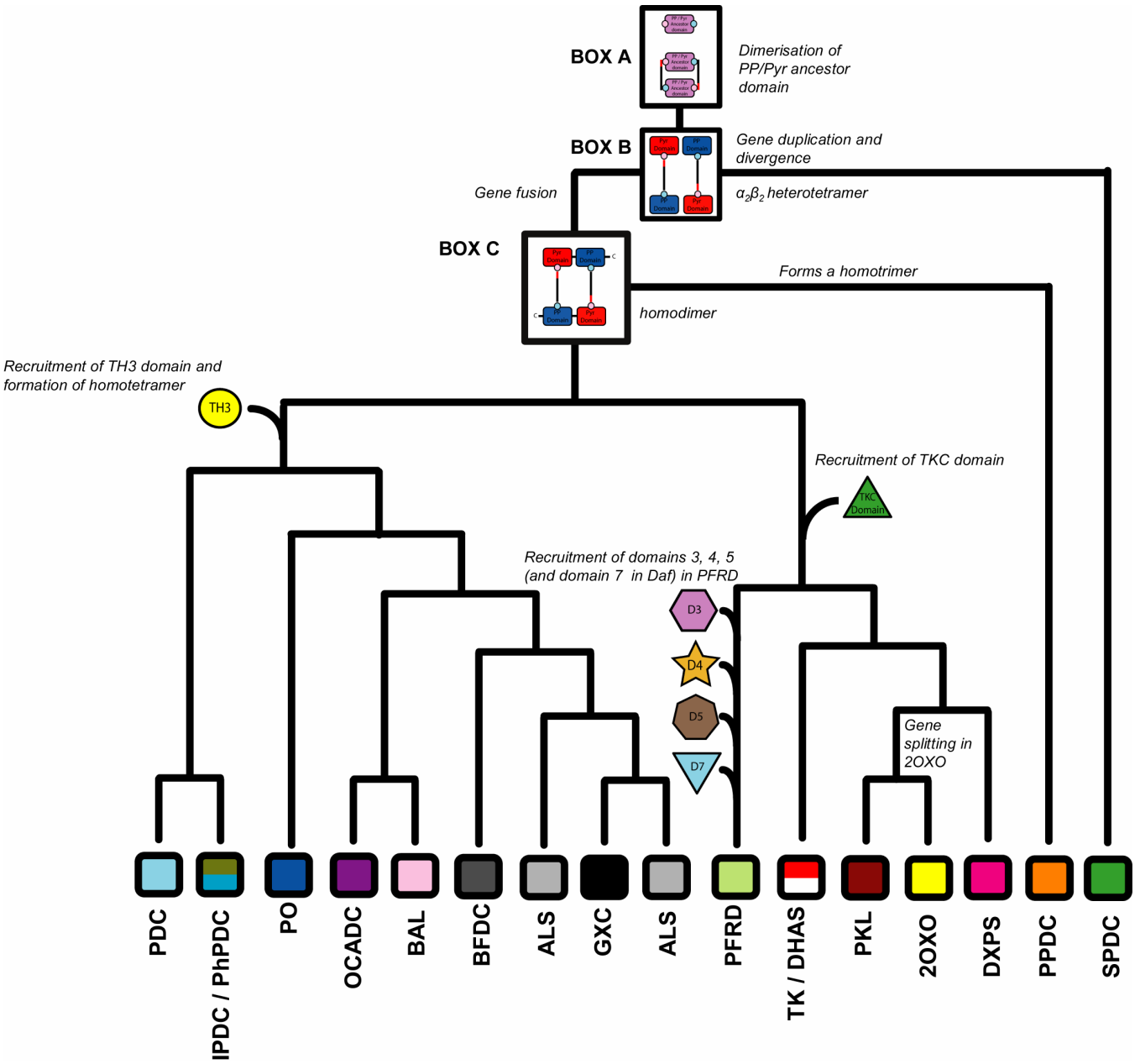

Figure 7: Proposed evolutionary history of the TPP-dependent enzyme family.

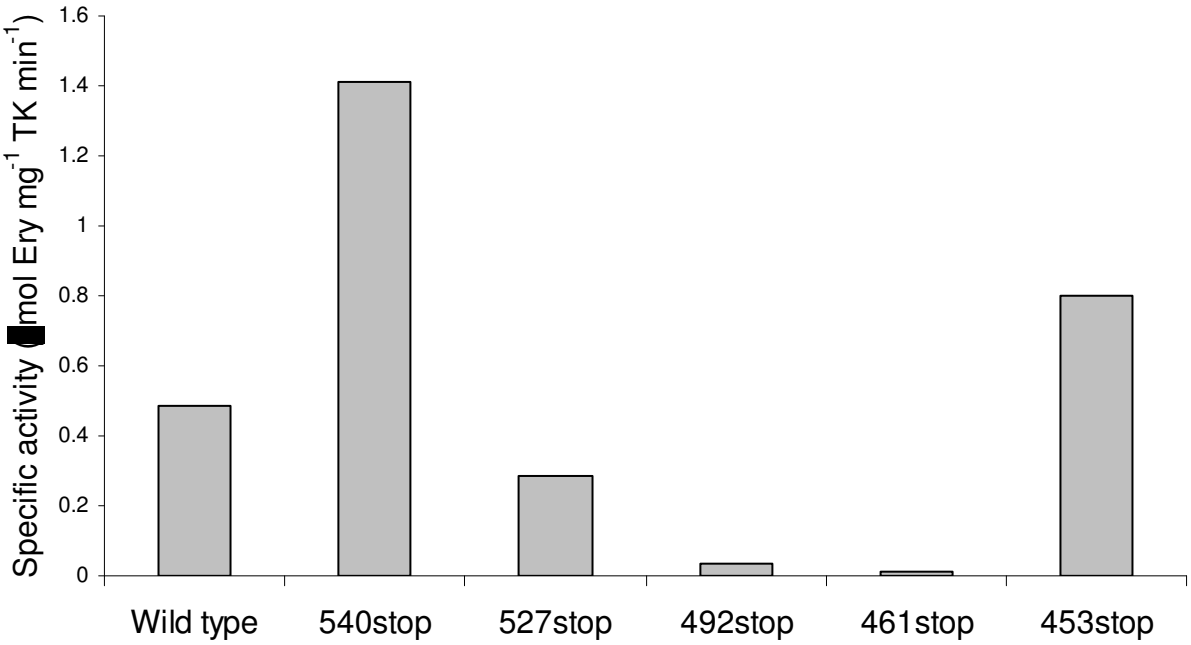

Figure 8: Specific activities of wild-type *Eco*TK and truncation mutants. TK catalysed reactions between 50 mM GA and 50 mM β-HPA in 50 mM Tris-HCl, 9 mM $MgCl_2$, 2.4 mM TPP, 0.5 mg ml^{-1} enzyme, pH 7.0, 25 ºC.

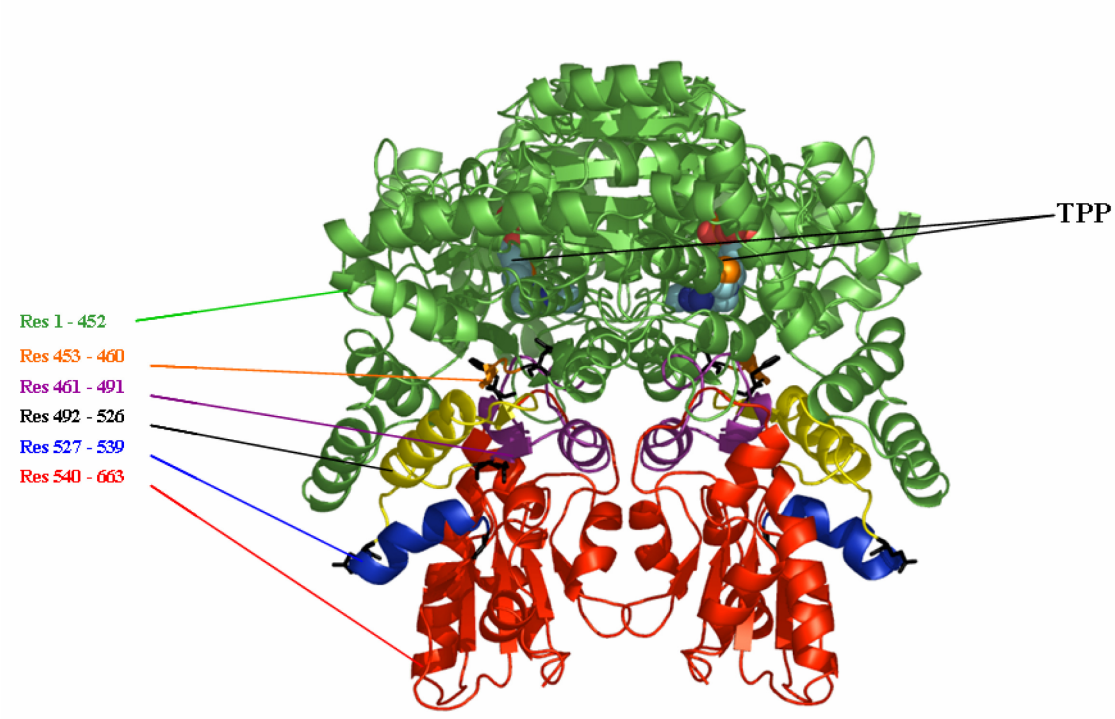

Figure 9: Location of C-terminal truncation mutants. Each coloured region, corresponding to the residue numbers shown, was removed consecutively from the enzyme C-terminus by insertion of stop codons into the TK gene. The region remaining in the smallest TK truncation variant is coloured green. The TPP cofactors are shown as spheres. Image generated using the 1QGD.pdb file in *Pymol.*

Eco TK	Residue	Function	% Cons. TK	DXPS	DHAS	PKL	2OXO	PFRD	PDC	IPDC	PhPDC	PO	ALS	GXC
26	**His**	Catalysis and stereospecificity	98	100	100	100								
66	**His**	Substrate recognition and binding	100	100	100	100								
100	**His**	Substrate recognition and binding	98		100	100								
154	**Gly**	H-bonds with diphosphate of TPP	94	100	67	100	100	100	100	100	100	100	100	100
155	**Asp**	Metal-binding	100	100	100	100	100	100	100	100	100	100	100	100
162	**Glu**	H-bonding network	96		100		77	94						
167	**Glu**	H-bonding network	98	100	100		23	29	100	100	100	100	93	100
187	**Asn**	Metal-binding	100	100	100	100	100	94	96	93	100	100	64	100
189	**Ile**	Metal-binding	100	91	33	100	23							
261	**His**	Catalysis and stereospecificity	96	8										
358	**Arg**	Phosphate binding	96											
381	**Asp**	Interacts with MT ring of TPP	96		33	7	100							
385	**Ser**	Phosphate binding	94			3		94				8		
411	**Glu**	Catalysis	100	100	100	100	100	100	100	100	100	100	100	100
434	**Phe**	Forms hydrophobic pocket	96	100	100	100			100	100	100			
437	**Phe**	Forms hydrophobic pocket	96	100	100		100							
440	**Tyr**	Forms hydrophobic pocket	96		100		100			20				
461	**His**	Phosphate binding	96		100	55								
469	**Asp**	Stereospecificity	96	100	67				4	27		38	36	
473	**His**	Transition state stabilisation	98	100	100	96	100	100	100	100	100		50	
520	**Arg**	Phosphate binding	98	100	67	96	100			27	67	4	47	

Sce PDC	Residue	Function	% Cons. PDC	TK	DXPS	DHAS	PKL	2OXO	PFRD	IPDC	PhPDC	PO	ALS	GXC
474	**Tyr**	TPP-binding	100							93	100	17	53	
477	**Glu**	Metal binding and catalysis	100							73	33			

Lpl PO	*Residue*	Function	% Cons. PO	TK	DXPS	DHAS	PKL	2OXO	PFRD	PDC	IPDC	PhPDC	ALS	GXC
380	**Val**	Substrate specificity	96	100	100	100	86			100	93	100	93	100

Daf PFRD	Residue	Function	% Cons. PFRD	TK	DXPS	DHAS	PKL	2OXO	PDC	IPDC	PhPDC	PO	ALS	GXC
991	**Thr**	Metal binding	100											
993	**Val**	Metal binding	100	4	100		97					100		
994	**Tyr**	Metal binding	100				3				20	58		

Ppu 2OXO	Residue	Function	% Cons. 2OXO	TK	DXPS	DHAS	PKL	PFRD	PDC	IPDC	PhPDC	PO	ALS	GXC
244	**Tyr**	Metal binding	100											

Table 1: A summary of the important residues in TK and other TPP-dependent enzymes examined. The function column and residue numbering refers to the enzyme for which a known residue function was first demonstrated or proposed. The frequency of occurrence of residues at equivalent positions in each enzyme is tabulated from the PP- and Pyr- alignments. Blank boxes indicate that no structurally equivalent residue was present.