Molecular cloning and Biochemical characterization of pyridoxal 5'-phosphate dependent enzymes of unknown function

Ph.D. Course in Biochemistry and Molecular Biology
XX Cycle, 2004-2007

Stefano Donini
Department of Biochemistry and Molecular Biology
University of Parma, Italy

Ph.D. program coordinator: Professor Gian Luigi Rossi
Department of Biochemistry and Molecular Biology,
University of Parma, Italy

Supervisor: Professor Alessio Peracchi
Department of Biochemistry and Molecular Biology,
University of Parma, Italy

General introduction

The main topic of this Ph.D. thesis is the characterization of gene products identified as PLP-dependent enzymes with unknown function. A complete and detailed description on function and evolution of PLP-dependent enzymes is provided in chapter 1.

Chapter 2 describes the molecular cloning as well as the recombinant expression and the initial biochemical characterization of a threonine synthase homolog from mouse. Finally, in chapter 3 I describe the study of two recombinant human proteins homolog of alanine-glyoxylate amino transferase 2.

The last part, Appendix, reports the evaluation of small RNA-cleaving DNAs (ribozymes), chemically modified by means of introduction of monomers of locked nucleic acids (LNA), in order to improve to cleavage of long and structured mRNAs.

Introduzione generale

Argomento principale di questa tesi di dottorato è la caratterizzazione di prodotti genici identificati come enzimi PLP-dipendenti, ma aventi funzione ignota. Nel capitolo 1 viene data una introduzione completa e dettagliata sull'evoluzione e le caratteristiche degli enzimi PLP-dipendenti.

Nel capitolo 2 sono descritti il clonaggio, l'espressione in forma ricombinante e l'iniziale caratterizzazione biochimica, di una protiena di topo omologa della treonina sintasi microbica. Nel capitolo 3 è illustrato lo studio di due proteine umane, omologhe dell'alanina-gliossilato amino transferasi 2.

Nell'ultima parte, l'Appendice, sono descritti studi su piccoli DNA ad attività ribonucleasica (ribozimi) modificati attraverso l'introduzione di monomeri di locked nucleic acid (LNA), con lo scopo di migliorare il taglio di mRNA dotati di struttura secondaria.

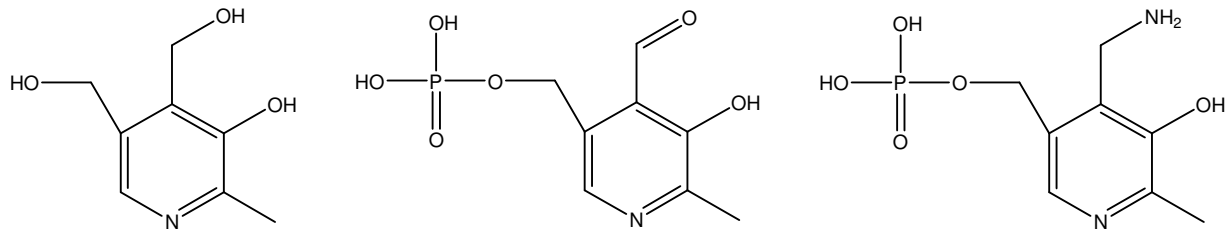
Contents

Chapter 1

Introduction

Summary

Following its identification (Heyl, 1951) vitamin B6, in the form of its biologically active phosphorylated derivatives pyridoxal-5'-phosphate and pyridoxamine-5'-phosphate (**Figure 1**), has been subject of extensive research since it represents the nature's most versatile cofactor (Vederas,1980; John, 1995; Jansonius, 1998; Metha & Christen, 2000; Schneider, 2000; Christen & Metha, 2001).



B6 **PLP** **PMP**

Figure 1. Chemical structures of vitamin B6 (**B6**), pyridoxal-5'-phosphate (**PLP**), pyridoxamine-5'-phosphate (**PMP**).

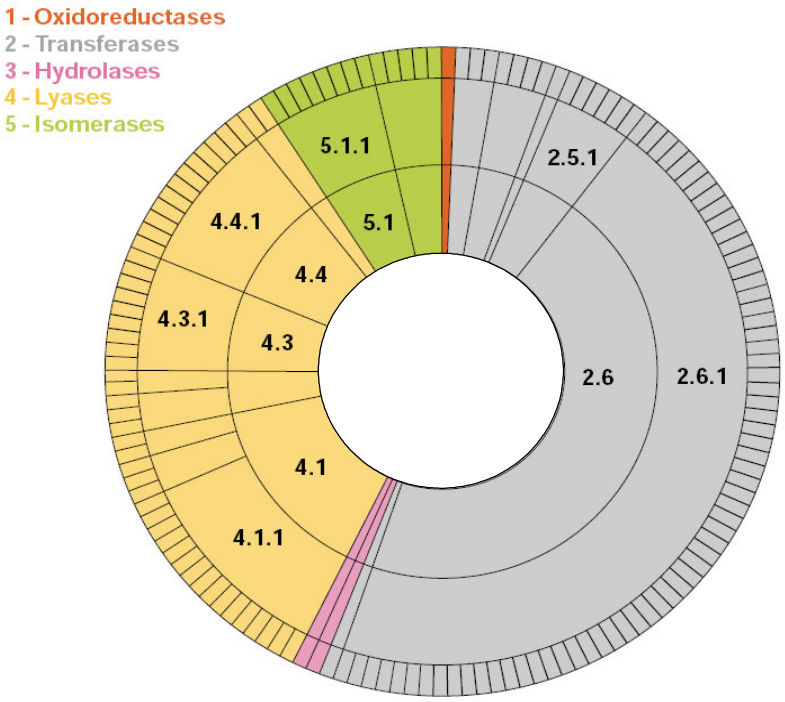
In fact, enzymes that use as cofactor pyridoxal-5'-phosphate (also called PLP-dependent enzymes) are unrivalled in the diversity of reactions that they catalyze, and show a widespread involvement in cellular processes. Versatility arises from the ability of pyridoxal-5'-phosphate to covalently bind the substrate and then to function as an electrophilic catalyst, thereby stabilizing different types of carbanionic reaction intermediates (Vederas, 1980; John, 1995; Schneider, 2000).

The functional diversity of PLP-dependent enzymes is illustrated by the fact that more than 140 distinct enzymatic activities catalogued by the Enzyme Commission (EC) are PLP dependent, corresponding to about 4% of all classified activities (**Figure 2**).

Almost all of these enzymes, with the exception of glycogen phosphorylase (Palm, 1990), are associated with biochemical pathways that involve amino compounds, mainly amino acids. Generally the reactions carried out by these enzymes that act on amino acids include transfer of the amino group, decarboxylation, interconversion of L- and D-amino acids, and elimination or replacement of chemical groups bound at the β- or γ-carbon.

Thanks to the amount of genomic information that has accumulated over the past few years, new genes have been identified as encoding PLP-dependent

enzymes, but a substantial fraction of these putative PLP-dependent enzymes remain functionally unclassified or only tentatively classified. For example, about of 20% of the putative PLP-dependent enzymes, encoded by the human genome, has not been assigned to an activity; moreover, more than one-third of all PLP-dependent enzymes classified by the EC are still uncharacterized (Christen & Metha, 2001; Percudani & Peracchi, 2003). Therefore, the characterization of putative PLP-dependent enzymes, together with the discovery of proteins for the "orphan" activities, represent fundamental challenges for the enzymologists, ensuing from the recent flood of genomic sequences.



Percudani & Peracchi, EMBO Report, Vol.4, No. 9

Figure 2. The catalytic versatility of the PLP-dependent activities classified is illustrated by a set of concentric pie charts (Thornton, 2000). Sector are coloured according to EC classes (1, oxoreductases; 2, transferases; 3, hydrolases; 4, lyases; 5, isomerases; 6, synthases). Of six general classes, five include pyridoxal-5'-phosphate (PLP)-dependent enzymes. The circles, from inner to outer, represent the first, second, third and fourth levels in the EC hierarchy. The angle substended by any segment is proportional to the number of activities it contains, making it clear that aminotransferases (EC 2.6.1) constituite the largest group of PLP-dependent enzymes.

My thesis work has focused on three putative PLP-dependent enzymes. Here, I will describe the preliminary bioinformatic analysis, the cloning of the cDNA encoding these proteins and the initial characterization of the purified recombinant proteins, with the goal to clarify their real activity.

Catalytic mechanism of PLP-dependent enzymes

A key feature that has emerged from the investigations of PLP-dependent enzymes is the versatility of the coenzyme in supporting very distinct catalytic actions. Although these enzymes are able to perform a host of different reactions on amino acids, there is a simple unifying principle explaining their mechanistic versatility (**Figure 3**). The first and common step for all PLP-dependent enzyme-catalyzed reactions is a Schiff base exchange reaction (transimination). Nearly all known PLP enzymes exist in their resting state as a "internal" aldimine, in which PLP is bound covalently via an imine bond to the ε-amino group of an active-site lysine residue.

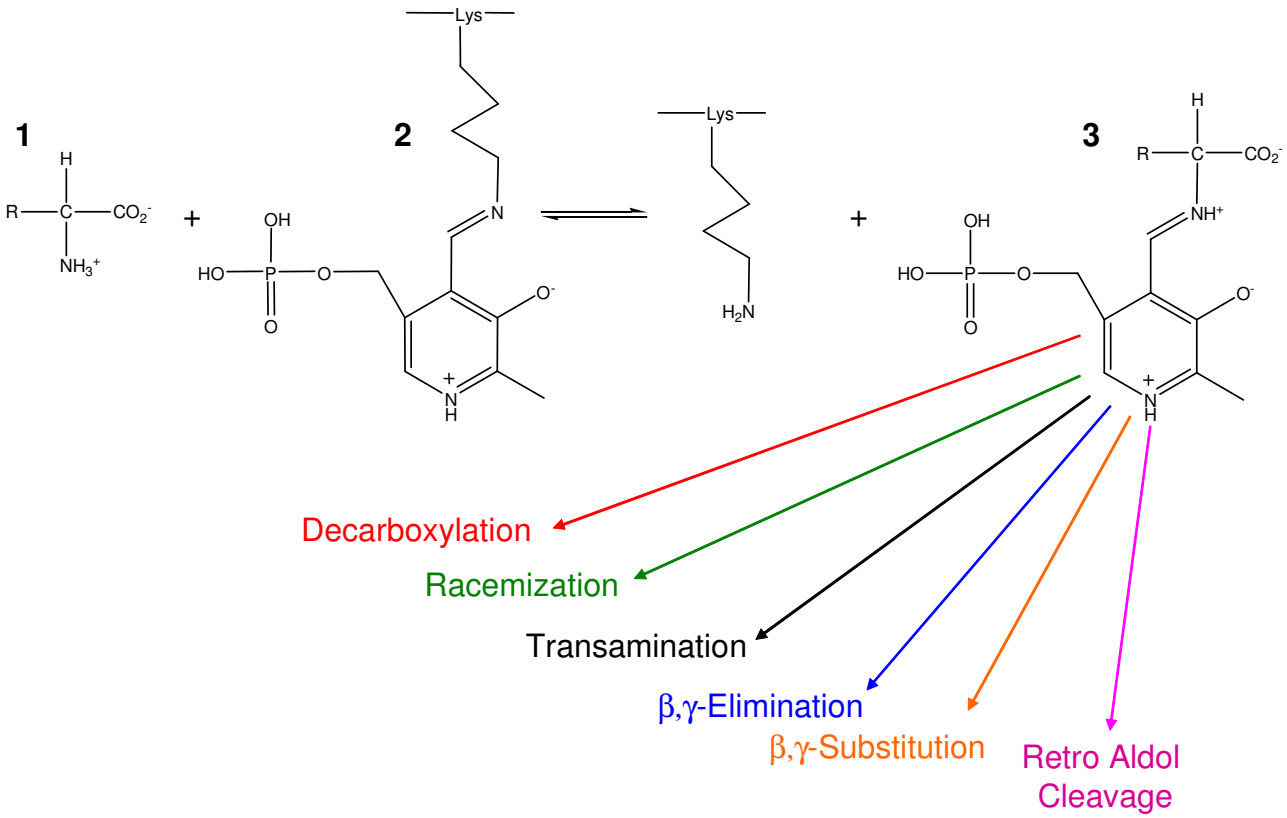


Figure 3. Transformations of amino acids by PLP-dependent enzymes. (**1**) Amino acid substrate. (**2**) "Internal" aldimine intermediate in which PLP has formed a Schiff base with the ε-amino group of an active-site lysine residue. (**3**) "External" aldimine intermediate. Thereafter, the reactions diverge, as shown . Formation of the planar pyridine-aldimine Schiff base **3** is a prerequisite for the catalytic efficacy of PLP, which is due to the electron-withdrawing effect exerted on Cα by the positively charged imine group and pyridine nitrogen atom and mediated through the extensive resonance system of the pyridine ring and the imine double bond.

In the first covalency change, the amino group of the incoming substrate replaces the lysine ε-amino group, forming a coenzyme-substrate Schiff Base

(external aldimine). This is a multiphasic process that includes several facile steps and is frequently very rapid compared to the central steps in the reaction mechanism (e.g., deprotonation of the substrate) (**Figure 4**). This "external" aldimine is the common central intermediate for all PLP-catalyzed reactions, enzymatic and nonenzymatic. Divergence in reaction specificity occurs from this point onward: depending on which bond at Cα is cleaved, one of the various reactions catalyzed by PLP-dependent enzymes can occur. Rupture of the covalent bonds involving Cα is possible because the coenzyme acts as an electron sink, stabilizing the negative charge development at Cα (carbanion) in the transition state, by delocalization through the pi system of the cofactor (Metha & Christen, 2000). In this way the electrons are stored from cleaved substrate bonds and subsequently dispensed for the formation of new linkages with incoming protons or second substrates. The completely formed carbanion is named quinonoid intermediate. This quinonoid resonance structure is commonly considered the major species responsible for the catalytic power of PLP, since the electrons from Cα are neutralized by the protonated pyridine nitrogen. In conclusion, the electron sink properties of pyridoxal 5'-phopshate are exploited to endow the holoenzyme with a variety of reaction types.

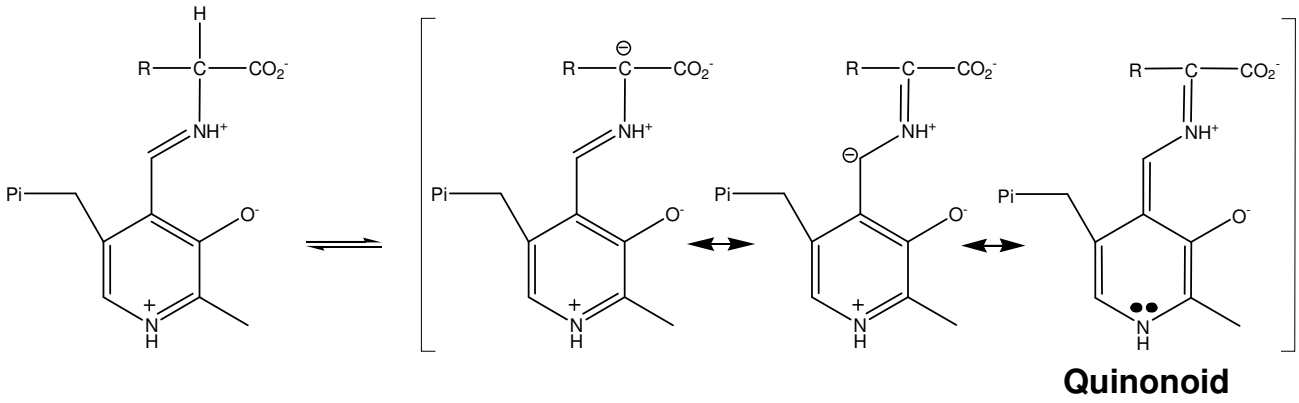


Figure 4. Principal resonance forms obtained from deprotonation of an aldimine intermediate. The quinonoid resonance form is labeled.

The multitude of reactions catalyzed by PLP-dependent enzymes can be divided according to the position at which the net reaction occurs. Reactions at the α position include transamination, decarboxylation, racemization, and elimination and replacement of an electrophilic R group (**Figure 5**).

Those at the β or γ position include elimination or replacement (**Figure 6**). Exceptions to these common types include the formation of a cyclopropane ring from S-adenosyl-L-methionine (SAM), catalyzed by 1-aminocyclopropane-1-carboxylate

(ACC) synthase (Adams, 1979), and the cleavage of ACC to α-ketobutyrate and ammonia, catalyzed by ACC deaminase (Walsh,1981).

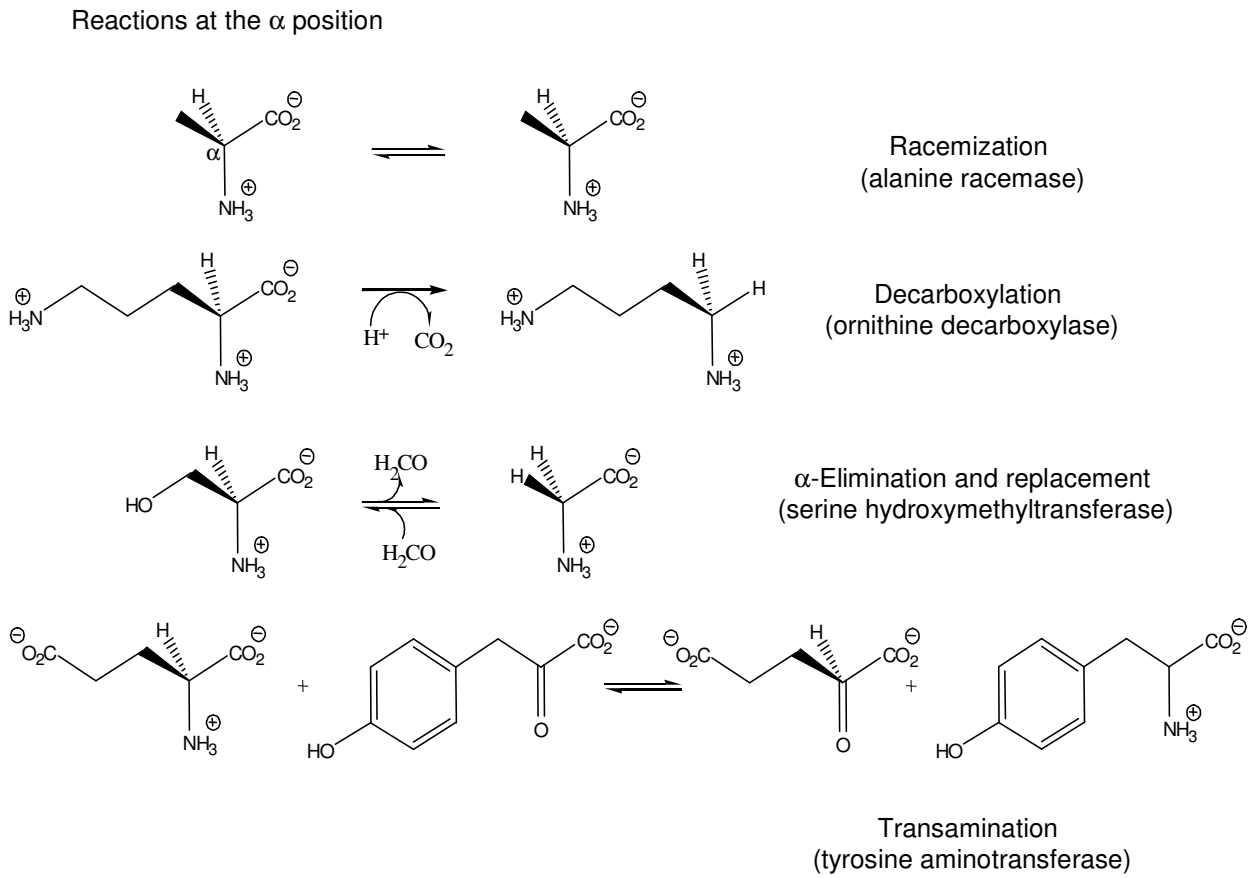


Figure 5. Examples of mechanism of reactions involving net change at the α position. All reactions are shown starting with the substrate aldimine, which is formed by transaldimination of the lysine-bound PLP. Racemization: the bacterial alanine racemase utilizes a tyrosine residue (Shaw, 1997; Sun, 1999) to deprotonate L-alanine, forming the quinonoid intermediate, which is reprotonated by a lysine residue on the opposite face of the cofactor to produce D-alanine. Decarboxylation: the reaction begins with loss of CO_2 from the substrate aldimine, producing the quinonoid intermediate. Protonation by an unidentified active site residue in ornithine decarboxylase generates the product aldimine. α-Replacement: the well-studied serine hydroxymethyltransferase (Rao, 2000) initiates the retro-aldol cleavage of serine by deprotonation of the hydroxyl group. Formaldehyde is released to generate the quinonoid intermediate. Protonation of the quinonoid at Cα by the lysine produces the aldimine of the product glycine. Transamination: the first half-reaction catalyzed by tyrosine aminotransferase involves initial proton abstraction from the glutamate aldimine Cα by the active site lysine, yielding the quinonoid intermediate. Reprotonation at C4' of the cofactor by that lysine generates the ketimine intermediate, which is subsequently hydrolyzed to release α-ketoglutarate, leaving the enzyme in PMP form. A complete catalytic cycle involves subsequent reaction with hydroxyphenylpyruvate to give tyrosine and to regenerate the PLP form of the enzyme.

Due to their common mechanistic features, PLP-dependent enzyme often show "catalytic promiscuity", that is, the ability of a single enzyme to catalyze different chemical reactions (O'Brein & Herschlag, 1999; Jeffery, 1999).

For example, L-serine racemase catalyses the deamination of L-serine at a rate similar to that of serine racemization (Strisovoky, 2003).

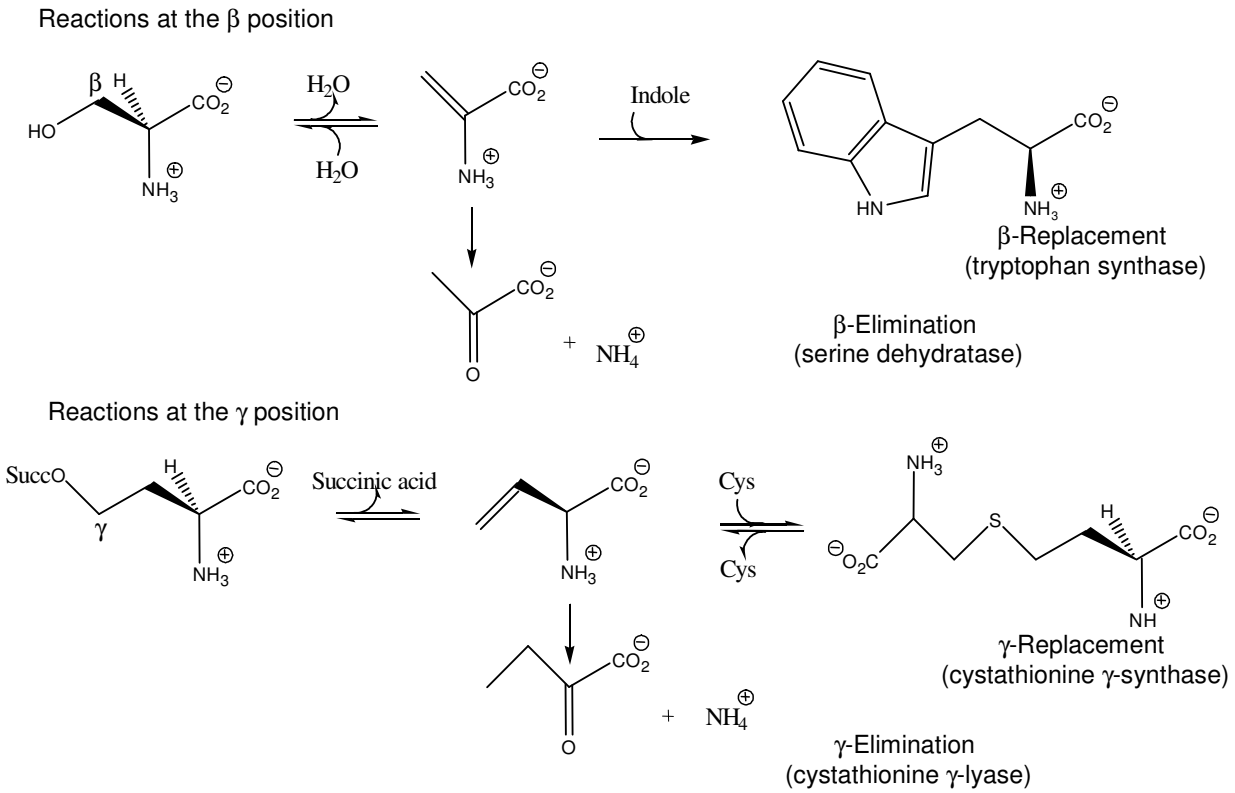


Figure 6. Examples of mechanism of reactions at the β and γ positions. β-Replacement: tryptophan synthase (Miles, 2001) catalyzes a net β-replacement by first deprotonating the serine aldimine at Cα, producing the quinonoid intermediate. Protonation of the hydroxyl group by the active site lysine promotes its elimination, generating the aldimine of amino acrylic acid. Indole adds to Cβ to form a second quinonoid that is subsequently protonated at Cα to generate the product aldimine. γ-Replacement: in the cystathionine γ-synthase-catalyzed γ-replacement reaction (Brzovic, 1990), *O*-succynil homoserine is deprotonated at Cα to produce the quinonoid intermediate that is subsequently protonated at C4' of the cofactor to give the ketimine intermediate. Proton abstraction at Cβ by an unknown active site base results in elimination for the succinyl group, which may occur in either a step-wise or concerted (shown) manner. Michael addition of cysteine to the β,γ-unsatured ketimine and subsequent proton transfers yield a second quinonoid intermediate that is protonated at Cα to form the product aldimine.

In addition, PLP-dependent enzymes may have non-strict substrate specificity (Han, 2001). A good example is threonine synthase (EC 4.2.3.1), the last enzyme of the threonine biosynthetic pathway. This enzyme catalyzes L-threonine formation from L-homoserine phosphate and also from DL-vinylglycine (Laber, 1994). Furthermore, threonine synthase shows "catalytic promiscuity", since it can catalyze β-elimination reactions with L-serine, DL-3-chloroalanine, L-threonine, and L-allo-threonine as substrates to yield pyruvate or alpha-ketobutyrate, while L-alanine, L-2-

aminobutanoic acid, and L-2-amino-5-phosphonopentanoic acid are substrates for half-transamination reactions to yield the pyridoxamine form of the enzyme and the corresponding alpha-keto acid.

Reaction specificity in pyridoxal phosphate enzymes

All reactions catalyzed by PLP-dependent enzymes are assumed to take place, very slowly, with substrates and PLP in the absence of the apoenzyme. Considering that PLP enzyme-catalyzed reactions corresponding to the nonenzymatic ones were shown to follow the same basic chemical mechanism, the determinants of reaction type specificity have to be shared.

In 1966 Dunathan proposed that the topology of the amino acid aldimine determined which bond to Cα would be broken (Dunathan, 1966). He postulated that the bond to be broken in the substrate-cofactor complex has to be perpendicular to the pyridine ring of the cofactor in the transition state of the reaction (**Figure 7**). It is in this conformation that the breaking σ bond achieves maximal orbital overlap with the π system (Corey,1956). This suggest that stereoelectronic effects play a major role because enzymes can very specifically orient one bond such that the nascent p orbital is aligned with those of the conjugated Schiff base/pyridine ring π system. The bond so oriented would be selectively labilized.

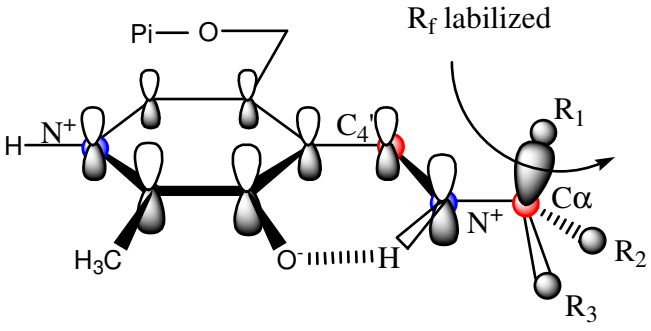


Figure 7. The Dunathan stereoelectric hypothesis. Substrates are bound to PLP such that the bond to Cα that is to be broken is aligned with the pi orbitals of the cofactor. Control of the substrates orientation thus enables the enzyme to distinguish among the reactions type.

In the absence of enzyme the rate of racemization and hydrogen exchange at the α carbon of amino acid-pyridoxal Schiff bases is determined by the proportion of aldimine substrate having the Cα-Hα bond orthogonal to the π system (Tsai, 1978).

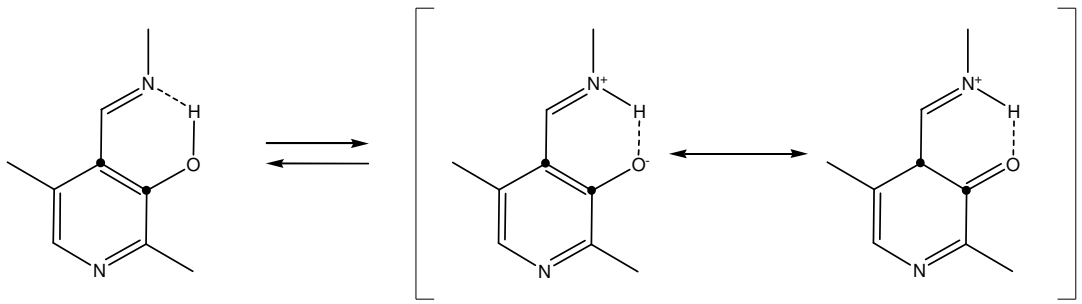
This model has been confirmed later, after the aspartate aminotransferase/phosphopyroxyl aspartate complex was solved (Kirsch, 1984). Ever since all the determined structures were in accord with Dunathan's hypothesis. Thus, the ensuing carbanion is stabilized by conjugation with the extended π system.

The role of the enzyme active site residues is to provide a rate enhancement and specificity which is unattainable in their absence. Certainly their major functions must include increased acid-base catalysis (Rios, 2000), anchoring of the phosphate group (Denessiouk, 1999), and recognition of the correct reaction partners (John, 1995), in order to enforce catalytic power as well as enhance selectivity of substrate binding, determining which of the many potential pathways is adopted by the coenzyme-substrate complex.

Spectroscopic properties of PLP-dependent enzymes

The absorption spectra PLP has been the subject of numerous studies (Metzler, 1980; Metzler, 1988). PLP shows in water a deep yellow colour and an absorption band at ~390 nm.

These features can be used to study the activity of a generic PLP-dependent enzyme. Usually, PLP-containing enzymes have still a deep yellow colour, but the absorption bands range from 412 to 435 nm. Furthermore, a minor band can appear at ~330-340 nm.



Enolimine tautomer
~330-340 nm

Ketoenamine form
~413-435 nm

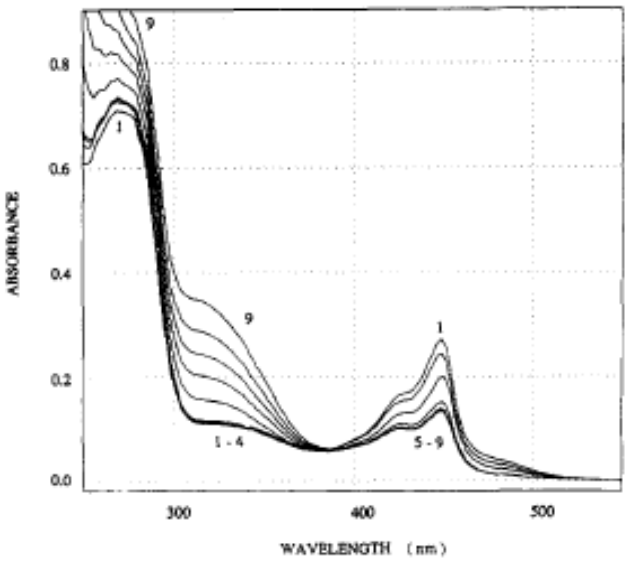
Figure 8. Tautomerism of Schiff bases of PLP. The absorption peak is shown for each tautomer form.

These two bands can be explained by the fact that the Schiff bases of PLP can exist in several different tautomeric forms. The predominant resonance form of

those species, having an unprotonated phenolic oxygen and a protonated imino group, is the ketoenamine structure, that is in equilibrium with the corresponding enolimine tautomer (**Figure 8**).

Since the enzyme-bound cofactor as well as certain PLP-derived reaction intermediates are known to have distinctive UV-visible absorption maxima, their spectrophotometric detection is used as a probe for substrate identification. Moreover in some cases it's also possible to identifiy kinetically competent intermediates in the PLP-enzyme catalyzed reactions. For example, the substrate specificity of a threonine synthase from *E.coli*, has been evaluated by looking the changes in the absorption spectra of the native enzyme (**Figure 9**) (Laber, 1994).

In the absence of substrates the spectrum of threonine synthase (EC 4.2.99.2) showed absorbance maxima at 333 and 416 nm, beside the protein absorption at 279 nm. Spectra taken about 10s after mixing with homoserine phosphate showed a sharp increase in absorbance at 448 nm and a broad shoulder extending from 460 to 510 nm, typical of a quinonoid intermediate.



Laber *et al.*, Biochemistry 1994, Vol.30, No.11

Figure 9. Absorbance spectra for the reaction of 20.0μM threonine synthase with 40mM L-homoserine phosphate in 50mM potassium phosphate buffer, pH 7.2, containing 0.15 M KCl. Spectra were collected 0.2 (l), 2.7 (2), 4 (3), 8 (4), 120 (S), 240 (6), 360 (7), 495 (8), and 675 (9) min after mixing.

Furthermore, it was possible to follow directly the formation of the product, because both maxima decrease in intensity with time, and after the absorbance at 448 nm finally reached a constant value, the absorbance in the region of 300-350 nm and below 280 nm, which is characteristic for of α-keto acids, started to increase slowly.

This is one of the multitude of examples in literature that capitalize the spectroscopic properties of the cofactor to study PLP-dependent enzymes.

Role of the closed conformation

As it happens with many enzymes not containing PLP as cofactor, an important features of a PLP-dependent enzyme, is the possibility of conformational changes after the substrate binding, or "induced fit". The contribution of induced fit to enzyme specificity has been much debated and refined over the years, albeit with a remarkable absence of experimental data.

The "induced fit theory" provides a mechanism for substrates discrimination, whereby, a good substrate induces an enzyme to adopt an active conformation, whereas a poor substrate cannot facilitate this conformational change (Koshland, 1958).

Now it is accepted that a conformational change induced only by certain substrates does not necessarily result in increased specificity for those substrates.

Thus, it has been argued that the conformational change observed in aspartate aminotransferase and many other PLP enzymes does not contribute to their substrate specificities (Hayashi, 1995).[¥]

Since it has been shown that release of the product oxoalacetate is partially rate determining for the reaction of aspartate aminotransferase with aspartate and α-ketoglutarate (Goldberg, 1996), aspartate aminotransferase is an example of an induced fit enzyme in which the conformational change may contribute to substrate specificity. It is not yet clear whether this conclusion extends to other PLP enzyme.

In contrast to the debated potential contribution of induced fit to substrate specificity (Pasternak, 2001), the closure of an active site upon substrate binding can certainly participate to reaction type specificity.

[¥]Exceptions to the general rule, however, include cases where substrate association or product dissociation is rate-determining for good substrates, whereas chemistry is rate-determining for poor substrate (Herschlag, 1988). In addition, the induced fit can decrease specificity if the two substrates induce different active conformations (Post, 1995).

Wolfenden pointed out the advantages of a closed conformation that makes more contacts with the bound substrate than are possible in a conformation from which the substrate can to dissociate (Wolfenden, 1974). Thus, greater control over proton transfer and solvent accessibility provided by the closed conformation can favor specific reactions.

A well-studied example is serine hydroxymethyltransferase, a PLP-dependent enzyme which catalyzes a variety of side reactions (transamination, decarboxylation, racemization, etc.) when presented with substrates other than serine.

The open form of the enzyme is unable to discriminate between the various reaction types in the way that the closed form can (Schirch, 1991). Therefore, these reactions occur, at least in part, because the substrates in question do not induce the closed conformation.

Dual specificity of aminotransferase

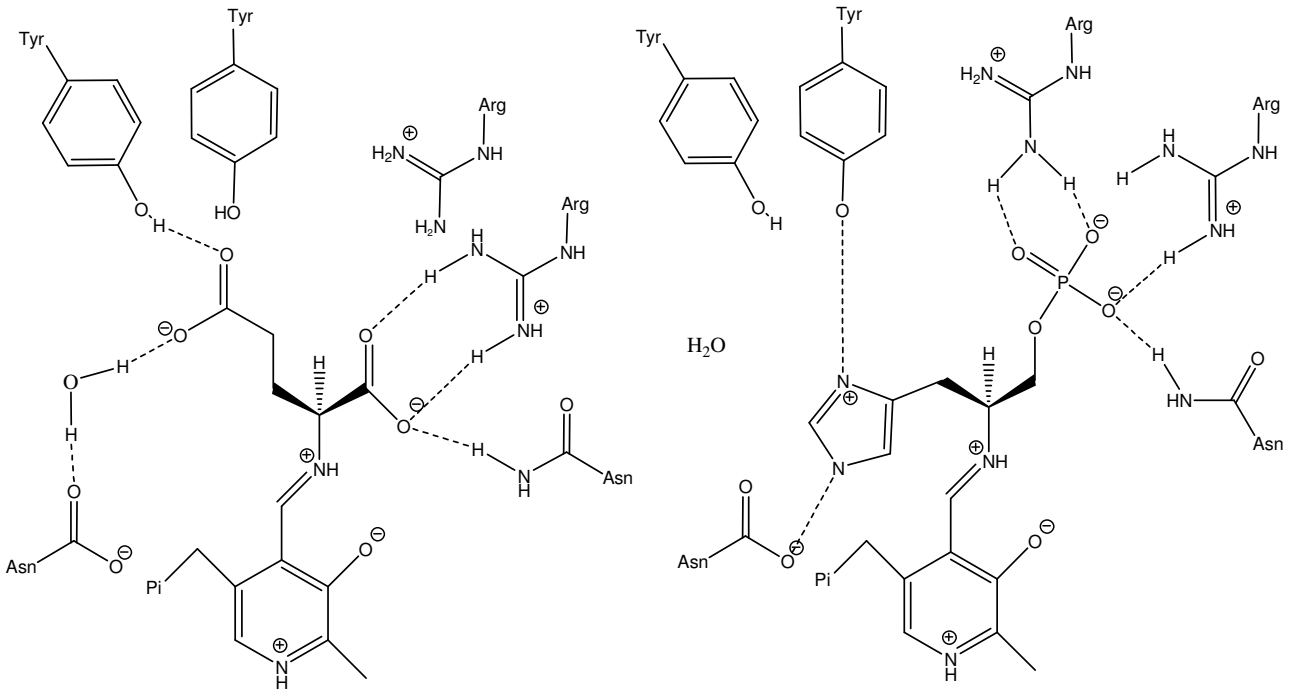
Because the aminotransferase reaction requires two different substrates to bind in succession to the same cofactor in the active site, these enzymes must be able to accommodate both structures while discriminating against all others.

One possible solution would be for the PLP itself to move between two different substrate binding sites, but such movement has never been observed.

An alternative is to take advantage of flexible side chains to position the functional groups into exclusive binding sites. This is the case for histidinol phosphate aminotransferase, which reacts with both histidinol phosphate and glutamate. The structures of substrate complexes of this enzyme (Haruyama, 2001) show that, although the phosphate and carboxylate groups interact primarily with the same arginine residue, the side chains of the two substrates have separate binding sites (**Figure 10**).

Most known aminotransferases, however, employ strategies for binding both substrates in the same site. Because aminotransferases many use the common substrate glutamate (thereby linking other amino acids to the cellular nitrogen pool), the problem of dual specificity is generally that of accommodating the negatively charged γ-carboxylate of glutamate in a site that must also accept a neutral or positively charged side chain.

Two common solutions to this problem have been found: the use of an arginine switch and an extended hydrogen bond network.



Glutamate aldimine Histidinol phosphate aldimine

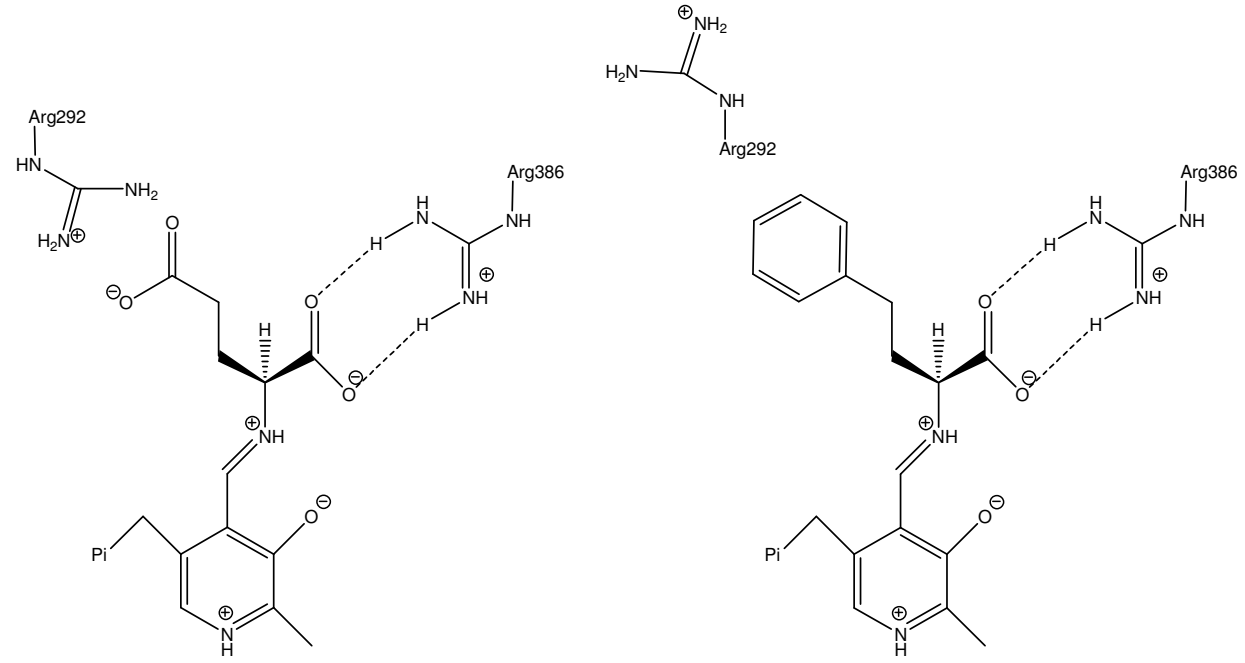
Figure 10. Dual specificity of histidinol phosphate aminotransferase. The phosphate of histidinol phosphate and the 1-carboxylate group of glutamate bind in the same site, although and additional arginine does interact with the phosphate group. In contrast, the oppositely charged imidazole and carboxylate side chains occupy spatially distinct sites and interact with different protein residues. The enzyme thus recognize each substrate specifically.

Arginine switches

The first example of an arginine switch was observed in an engineered enzyme that was constructed by introducing six mutations into aspartate aminotransferase. These changes substantially increased activity toward aromatic substrates (Onuffer, 1995). The subsequent determination of the structure of the mutant enzyme (Malashkevich, 1995) showed that the large aromatic substrates are accommodated by movement of Arg292 out of the active site. Dicarboxylate substrates bind their β- and γ-carboxylate by directly interacting with this arginine through a bidentate hydrogen bond/ion pair in the canonical aspartate aminotransferase (**Figure 11**). The position of Arg292 is locked in aspartate aminotransferase, thus substrates lacking a carboxylate side chain are effectively excluded.

More recently, direct evolution techniques have been used to broaden the substrate specificity of aspartate aminotransferase to include branched chain (Yano, 1998) or aromatic amino acids (Rothman, 2003). Crystal structures of mutants with increased

activity toward branched chain amino acids (Oue, 1999; Oue, 2000) indicated that they also have acquired the ability to switch Arg292 out of the active site when uncharged substrates are bound, indicating that this trait is easily induced in aspartate aminotransferase and is crucial for dual specificity.



Glutamate aldimine Phenylalanine aldimine

Figure 11. Schematic of the arginine switch in aminotransferases that react with both dicarboxylic and aromatic amino acids. The γ-carboxylate of glutamate (*left*) interacts closely with Arg292. This residues reorients to point out the active site when aromatic substrates bind (*right*). This movement allows the enzyme to accept both types of substrates.

Arginine switches are not unique to engineered enzymes. Tyrosine (aromatic) aminotransferase is a well-characterized enzyme that has natural specificity for the aromatic amino acids tyrosine, phenylalanine, and tryptophan, as well as for the dicarboxylic amino acids aspartate and glutamate.

A number of structures are now available of the *Paracoccus denitrificans* tyrosine aminotransferase that clearly demonstrate the arginine switch (Okamoto, 1998; Okamoto, 1999).

Crystallographic and modelling studies illustrated a similar strategy employed by the GABA (Toney, 1995; Storici, 1999) and ornithine (Storici, 1999) aminotransferase, which react with both ω-amino acid substrates, and with the common substrate glutamate. GABA aminotransferase, like aspartate aminotransferase and tyrosine aminotransferase, binds the dicarboxylic acid substrate via two conserved arginines. The carboxylate of the ω-

amino acid, GABA, occupies the same position as the γ-carboxylate of glutamate, thereby taking advantage of the similar distance between the amino and carboxylate groups of GABA and the amino and γ-carboxylate groups of glutamate. The second arginine (equivalent to Arg386 of aspartate aminotransferase) moves to interact with a conserved glutamate near the active site (**Figure 12**).

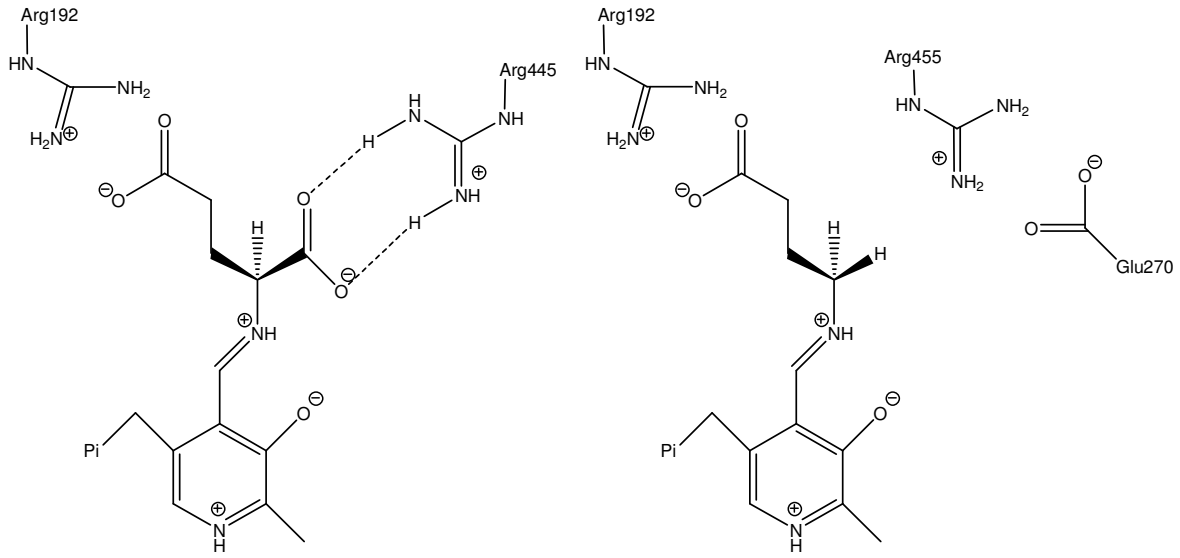
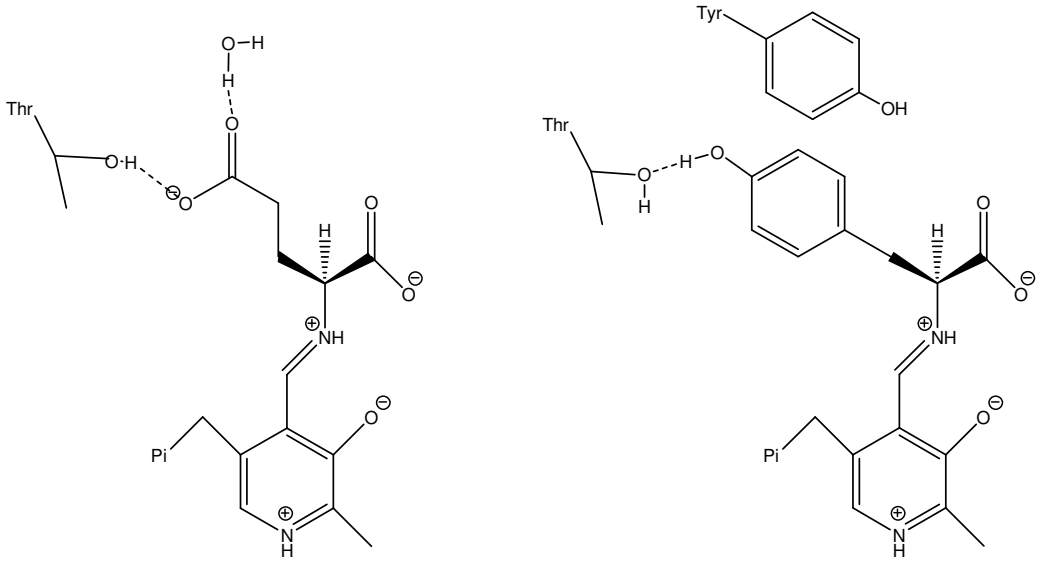


Figure 12. Schematic of the arginine switch in GABA aminotransferase. As in the case of Aspartate amino transferase, GABA aminotransferase binds the dicarboxylic acid substrate glutamate via two conserved arginines. In order to accommodate GABA, Arg445 moves away from the cofactor to engage in a salt bridge with a nearby glutamate residue.

Hydrogen bond networks

Arginine switches, however are not ubiquitous, even among tyrosine aminotransferases. Early sequence alignments indicated that a subgroup of aminotransferase lack an Arg292 equivalent (Jensen, 1996).

One of the structures available for a tyrosine aminotransferase of this group is that of the unliganded *Pyrococcus horikoshii* enzyme (Matsui, 2000). Modeling of the substrates in the active site suggests that this enzyme binds glutamate via an extended hydrogen-bonding network, as has been observed in the aspartate aminotransferase from this same organism (**Figure 13**) (Ura, 2001).



Glutamate aldimine Tyrosine aldimine

Figure 13. Dual substrate specificity can also be achieved by hydrogen bond rearrangement. *Pyrococcus horikoshii* TATase binds the γ-carboxylate of glutamate via hydrogen bond network rather than an arginine residue. Direct interactions are made with a threonine residue and a tightly bound water molecule. Modeling of bound tyrosine suggests that the hydrogen bond network rearranges, so that the aromatic ring stacks against a nearby tyrosine residue, as well as makes a hydrogen bond to the same threonine residue that is involved in glutamate association.

The absence of a positively charged residue in this tyrosine aminotransferase makes it much easier to accommodate the uncharged substrates by simple rearrangement of the hydrogen bond network.

The solved structures of the *E.coli* branched chain amino acid aminotransferase (BCAT) (Hayashi, 1993) show that this enzyme also takes advantage of different hydrogen bond interactions. BCAT reacts with the branched chain amino acids isoleucine, leucine, and valine, as well as with glutamate.

The hydrophobic substrates and glutamate all bind in the same site, which is a hydrophobic pocket consisting largely of aromatic residues.

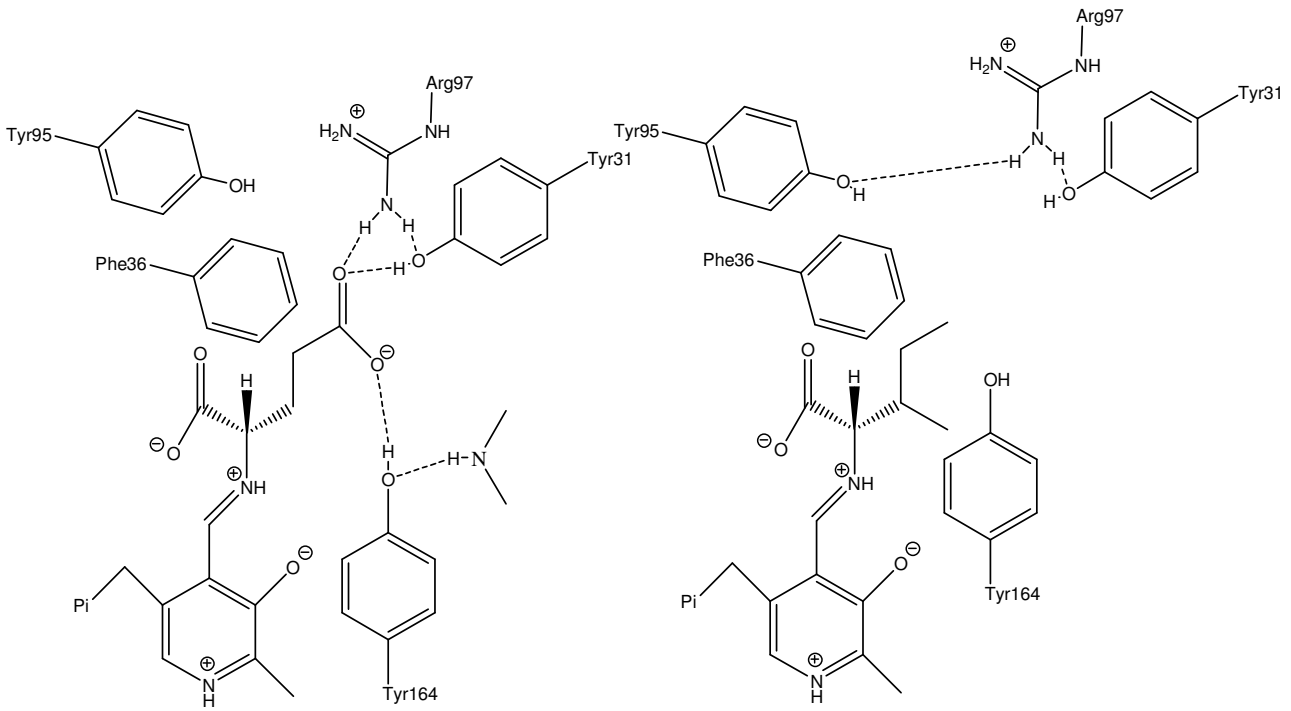


Figure 14. Dual substrate recognition by branched chain amino acid aminotransferase. The substrate binding pocket is composed primarily of aromatic residues, and the hydrophobic substrate isoleucine is surrounded by five of them, only three of which are shown (Phe36, Tyr164, and Tyr31). The longer glutamate substrate extends far enough to form hydrogen bonds with the hydroxyl groups of two tyrosines and the guanidinio group of an arginine residue.

In contrast to the tyrosine aminotransferase, there is no large–scale rearrangement of the H-bond network to accommodate the charged substrate. Instead, the glutamate side chain, which is slightly longer than that of the other substrates, protrudes into a pocket where four residues form hydrogen bonds with the carboxylate.

The arginine in this case forms with the substrate only a monodentate interaction rather than the bidentate H-bomd network seen in other enzymes, and it is extensively H-bonded to neighboring residues, allowing it to remain in position when the small hydrophobic substrates are bound (**Figure 14**).

Structural diversity in PLP-dependent enzymes

PLP enzymes were for a time underrepresented in protein structure databases (John, 1995), but the situation has changed in recent years, as an abundance of new structures have been solved. Although it was initially postulated that the structures of PLP enzymes would correlate with the reaction type (Alexander, 1994), it has since been realized that each of the major structural classes contains representatives of multiple reaction types. The fact that completely different folds are utilized as a scaffold for PLP catalysis indicates that enzymes based on this cofactor has emerged at several occasions during evolution.

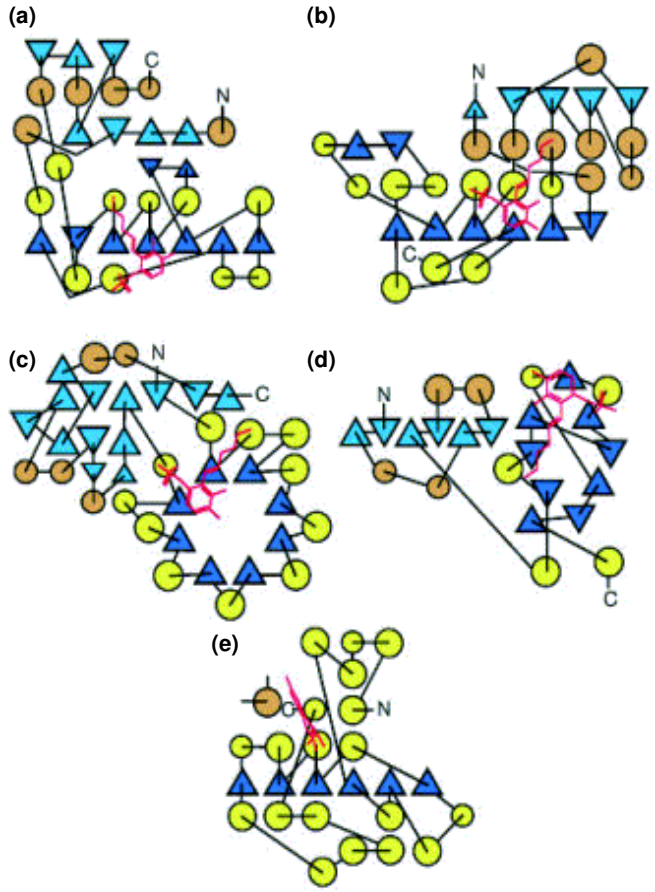


Figure 15. Examples of the five known pyridoxal-5′-phosphate (PLP) binding folds. Secondary structure elements are color coded: in the PLP-binding domains β strands are shown in dark blue and α helices in yellow; for other domains β sheets are in cyan and α helices in orange. PLP, bound covalently to the sidechain of the catalytic lysine residue, is shown in red. The topology diagrams are adapted from the Tops database. **(a)** Diamino-pelargonic acid synthase from *E. coli* (PDB code 1qj5), a representative of the fold-type I family. **(b)** Tryptophan synthase (PDB code 1ubs), fold-type II. **(c)** Alanine racemace (PDB code 1sft), fold-type III. **(d)** d-Alanine aminotransferase (PDB code 1daa), fold-type IV. **(e)** Glycogen phosphorylase (PDB code 1gpb), fold-type V; for clarity, only the C-terminal PLP-binding domain is shown. In addition, an α helix from the N-terminal domain (shown in orange) interacts with the pyridine ring of PLP and is included in the fold diagram.

The vast majority of PLP-dependent enzymes whose structures are available, can be subdivided into five independent fold-types (Grishin, 1995; Schneider, 2000) , that are, the fold-type I with aspartate aminotransferase as prototype enzyme (or aspartate aminotransferase family), the fold-type II with tryptophan synthase β as prototype (or tryptophan synthase β family), the fold-type III with alanine racemase family (or alanine racemase family), the fold-type IV with D-alanine aminotransferase as prototype (or D-alanine aminotransferase family), the fold-type V with glycogen phosphorylase (**Figure 15**).

While the five enzyme folds are unrelated with each other, a few features in the interactions of PLP with the active site are conserved. In all cases, the phosphate group is bound to the N terminus of an α helix, and a covalent Schiff-base linkage is used to anchor the cofactor to the enzyme in the resting state. In four of five folds, the catalytic lysine residue is located in the connection between a β strand and the following α helix.

Moreover, other new fold-types have been discovered: the fold-type VI and the fold-type VII with as prototype enzymes, respectively, the D-Lysine-5,6-aminomutase and the Lysine-2,3-aminomutase.

Fold-type I: The aspartate aminotransferase family

The majority of known structures belong to this fold-type, a group that includes the structurally best characterized PLP enzymes. The aspartate aminotransferase is not only the largest but also the functionally most diverse family.

The enzymes in this class invariably function as homodimers, but in some cases assemble into larger complexes. Each subunit folds into two domains, with two active sites per dimer.

Each monomer consists of a large domain, in which the central feature is a seven-stranded β sheet, and of small domain, comprising the C-terminal part of the chain, which folds into a three- or four-stranded β sheet covered with helices on one side.

The N-terminal part of the chain does not have a common fold within this class but often contributes to the small domain.

PLP is covalently attached to the large domain via the ε-amino group of a lysine residue at the N-terminus of a short helix following a β strand. The phosphate is anchored

to the N terminus of another a helix on the other side of the β sheet such that the aromatic ring of PLP packs against the neighbouring β strands.

The active site lies at on the dimer interface and residues from both domains and both subunits are involved in cofactor binding.

A multiple structural alignment including ten different enzymes within the family of fold-type I enzymes (Kack, 1999) showed that, in general terms, the subunits are of course similar within this structural family. Despite their similar folds, however, only five segments, corresponding to β strands β4, β7, β8, β9 and β10 in diamino pelargonic acid synthase, are strictly superimposable between these enzymes.

In addition, an analysis of the pairwise structurally aligned enzymes provided a basis for evaluation of evolutionary relationships within this fold-type. This analysis suggested that the fold-type I enzymes can be divided into six subclasses, of which three are aminotransferase subclasses (Grishin, 1995; Metha, 1993). This classification correlates with the structure of the N-terminal part of the chain, which folds similarity within a subclasses, but shows entirely different conformations between the subclasses (Kack, 1999).

Based on difference in the structure of the N-terminal part, two further subclasses can be defined: the serine hydroxymethyltransferase subclass and the 3-amino-5-hydroxybenzoic acid synthase subclass (Schneider, 2000).

Superposition of the fold-type I enzymes reveals that the location of the cofactor in the active site of these enzymes is virtually identical. However, this comparison suggests that the active-site lysine residues are not structurally conserved (Kack, 1999). In the subclass 2 aminotransferases this residue occurs one position earlier in the polypeptide chain than in the other fold-type I enzymes, suggesting a deletion in the ancestral subclass 2 aminotransferase. Instead, the only structurally and functionally conserved residue within this fold-type family is an aspartic acid residue that interacts with the pyridine nitrogen, indicating its importance for catalysis.

Enzyme-cofactor interactions of importance for the mechanism in aspartate aminotransferase, such as parallel stacking of the pyridine ring from the *re* side and the hydrogen-bond network to phenolic oxygen, are not conserved.

These discrepancies suggest that different means are used to tune the catalytic process in the different subclasses.

Fold-type II: The tryptophan synthase β family

This family includes among other enzymes the tryptophan synthase β subunit (Hyde, 1988), threonine deaminase (Gallagher, 1998) and Threonine synthase (Lambert, 1998).

Tryptophan synthase fold into an $\alpha_2\beta_2$ tetramer, where β is the catalytic PLP-binding subunit and α is a regulatory subunit. Threonine deaminase, active as a homodimer, similarity contains one regulatory domain and one PLP-binding, catalytic domain; however, these two domains are accommodated within one subunit of the enzyme.

The PLP-binding regions in these enzymes share the same fold and consist of two domains of about equal size: an N-terminal domain containing a four-stranded sheet surround by helices; and C-terminal domain built up by a six-stranded sheet with flanking helices.

The active site is accommodated in a cleft formed between the C-terminal ends of the two β sheets. PLP is bound to an active site lysine residue, which is located at the beginning of an α helix just after a strand in the N-terminal domain.

The phosphate group binds to the N terminus of another helix in the C-terminal domain, which positions the aromatic ring of PLP at a cross-over connection in this β sheet.

The cofactor is bound in the same orientation as in the aspartate aminotransferase (i.e., with the *re* face towards the solvent). In these enzymes, a serine residue coordinates the pyridine nitrogen, rather than the aspartic acid used in the aspartate aminotransferase family.

Fold-type III: The alanine racemase family

Alanine racemase (Shaw, 1997), eukaryotic ornithine decarboxylase (Kern,1999), and a putative PLP-dependent enzyme (Eswaramoorty, 1999) are the only members of this family with known three-dimensional structure. Other enzymes such as prokaryotic diaminopimelate decarboxylase, and biosynthetic arginine decarboxylase have been assigned to this class on the basis of similarities in primary sequence (Grishin, 1995).

Alanine racemase is a homodimer in which each subunit folds into two domains, an eight stranded α/β barrel and a domain mainly comprising β strands.

PLP binds at the mouth of the α/β barrel in a cleft formed between the two domains. The active-site lysine is located at the C-terminal end of the first strand where it turns into the proceeding helix.

The phosphate group of PLP interacts with the N terminus of the last helix of the α/β barrel. Contrary to the PLP-dependent enzyme classes described above, PLP binds with its *re* side facing the protein.

The sidechain of an arginine residue forms a hydrogen bond to the pyridine nitrogen, indicating that protonation of the pyridine nitrogen is not crucial for catalysis in these enzymes. An arginine residue at this position was also observed in a putative PLP-dependent enzyme identified in yeast (Eswaramoorty, 1999).

In the mammalian ornithine decarboxylases, however, the environment of the PLP resembles that in the aminotransferases, with a hydrogen-bond interaction to an acidic glutamic acid residue (Kern,1999).

Fold-type IV: The D-amino acid aminotransferase family

This class presently has only two members with known three-dimensional structure: D-amino acid aminotransferase (Sugio, 1995) and branched-chain aminotransferase (Okada, 1997). Goldsmith and colleagues (Grishin, 1995) also assigned 4-amino-4-deoxychorismate lyase to this class.

The fold-type IV enzymes are superficially similar to fold-type I and II. The fold consists of a two-domain structure with the active site located at the domain interface. Two identical subunits form a catalytically competent dimer; however, branched-chain aminotransferase further assembles into a hexamer.

The smaller N-termianl domain contains a six-stranded antiparallel β sheet with two α helices on one side. In the larger C-terminal domain, two four-stranded β sheets form a pseudo-β-barrel that is surrounded by a few helices.

The active-site lysine is situated in a loop between the first strand in the barrel and a following helix. The phosphate interacts with the N terminus of an α helix in the same domain.

The cofactor binds with its *re* side facing the protein rather than the active-site pocket as in the fold-type I family, accounting for the difference in stereochemistry of the products in the reaction of D-amino acid aminotransferase. A glutamic acid sidechain forms a hydrogen bond to the pyridine nitrogen.

The active site of D-amino acid aminotransferase is virtually a mirror image of the active site of the fold-type I aminotransferases, and these two enzyme families provide an excellent example of convergent evolution (Sugio, 1995).

Fold-type V: The Glycogen phosphorylase family

The crystal structure of glycogen phosphorylase was the first structure of a PLP-binding enzyme to be determined (Weber, 1978; Sprang & Fletterick, 1979). In members of this class of enzymes, PLP does not act as an electrophilic catalyst, but instead its phosphate group participates in proton transfer. However, because glycogen phosphorylase binds PLP in a very specific manner and PLP is used for catalytic purposes this enzyme can be included in the superfamily of vitamin B6 dependent enzymes.

Glycogen phosphorylase is a multidomain protein consisting of an N-terminal domain of approximately 310 residues, a glycogen-binding domain spanning 160 residues, and a C-terminal domain of about 360 residues. The C-terminal domain has a dinucleotide-binding fold and binds the cofactor PLP. The phosphate group interacts with the N terminus of an α helix that also harbours the active-site lysine residue one turn further down the helix. No hydrogen bonds are made from the protein to the pyrimidine nitrogen.

Two new fold-type families: D-Lysine-5,6-aminomutase & Lysine-2,3-aminomutase family

Recently have been identified two new PLP-dependent enzymes that do not fall into the five fold-type categories described above and do not resemble any of the known PLP-dependent enzymes, the D-Lysine-5,6-aminomutase (5,6-LAM; Berkovitch, 2004) and the Lysine-2,3-aminomutase (2,3-LAM; Lepore, 2005), representing respectively the Fold-type VI and Fold-type VII.

Fold-type VI: The D-Lysine-5,6-aminomutase family

The D-Lysine-5,6-aminomutase, it is an adenosylcobalamin and pyridoxal-5'-phosphate dependent enzyme that catalyzes a 1,2 rearrangement of the terminal amino group of DL-lysine and of L-β-lysine. 5,6-LAM cannot be placed into any of five PLP enzyme families, although it does share features with fold types II, III, and IV.

The overall structure of 5,6-LAM is an $\alpha_2\beta_2$ tetramer (Baker , 1973) that can be thought of as a dimer of $\alpha\beta$ units. The large α subunit (538 residues) is composed of the PLP-binding TIM barrel domain and several additional α-helices and β-strands at the N and C termini. The small β subunit (262 residues) comprises two domains: the N-terminal dimerization domain, which has the same fold as the Cu-binding domain of the Alzheimer's disease amyloid precursor protein (Protein Data Bank ID code 1OWT), and the AdoCbl-binding Rossmann domain, which also provides the imine bond to PLP. The Rossmann domain interacts with the C terminus of the TIM barrel, placing PLP into the top of the barrel while projecting AdoCbl to the edge of the barrel, far from the PLP-binding site.

The distinctive feature among all PLP-dependent enzymes for the D-Lysine-5,6-aminomutase is that the imine linkage is formed to the Rossmann domain rather than to a lysine of the TIM barrel.

Fold-type VII: The Lysine-2,3-aminomutase family

Lysine-2,3-aminomutase, the only members of this family, catalyzes the interconversion of L-α-lysine and L-β-lysine. It does not require a vitamin B12 coenzyme but instead requires a [4Fe-4S] cluster, *S*-adenosyl-L-methionine (SAM), and pyridoxal-5_-phosphate (PLP) as coenzymes (Petrovich, 1992; Frey, 2001).

Lysine-2,3-aminomutase crystallizes as a homotetramer, consisting of a dimer of tightly associated dimers. The subunits of the dimer are linked through a zinc ion that coordinates residues from both subunits.

Each subunit comprises three domains. The central globular domain composed is reminiscent of a TIM barrel $(\beta/\alpha)_8$; however, the barrel lacks two β-strands and forms a $(\beta/\alpha)_6$-crescent. The β-sheet of the crescent is extended by two β-strands at the C-terminal side, the last of which is antiparallel, and an antiparallel β-strand at the N-terminal side. An N-terminal domain comprises seven short helices forming a cluster closing off the C terminus of the channel formed by the central crescent of the opposing subunit. The C-terminal domain contains cysteine residues which coordinate the zinc ion, and a three stranded β-sheet that interacts with the last two β-strands of the crescent formed by the opposing subunit containing the fourth zinc ligand.

Thus, the quaternary structure involves an assembly of two domain-swapped dimers that are held together by zinc ions in addition to C-terminal coils that converge at the tetramer equatorial region in the formation of an extensively hydrogen-bonded tubular

structure. The C-terminal domain accounts entirely for the contacts in the dimer–dimer interface, the remaining contacts being mediated by water molecules.

A striking feature of the active site is the hydrogen-bonded contact between PLP and a fixed water molecule held in place by backbone hydrogen bonds from the carbonyl of Arg-116, amide hydrogen atoms from Tyr-113 and Arg-112, and N1 of PLP. The fixed water most likely donates a proton in the hydrogen bond to N1 of PLP, and the pyridine ring is not protonated. This mode of PLP binding is unprecedented among PLP-dependent enzymes.

Evolutionary relationship among PLP-dependent enzymes

PLP provides an impressive example of how biological evolution has exploited the catalytic potential of organic cofactors. The versatility of PLP as prosthetic group of enzymes is perhaps equaled by zinc as inorganic cofactor of all EC classes of enzymes (Vallee & Falchuk, 1993). The indispensable mechanistic role of PLP, the multiple evolutionary origin of the enzymes that depend on it, and the uniqueness of PLP as cofactor for the transamination of amino acids all argue for PLP to arrive on the evolutionary scene before the emergence of the apoenzymes. This hypothesis is confirmed by the recent recognition that the main metabolic pathway for the *de novo* biosynthesis of PLP, based on just two enzymes termed Pdx1 and Pdx2, is remarkably conserved in plants, fungi, archaea and most bacteria (Mittenhuber, 2001; Fitzpatrick , 2007). The conservation of this biosynthetic route among modern-day organism[¥] suggest, at the very least, that the pathway originated before the separation between the three kingdoms of life.

There are general arguments suggesting that organic cofactors and coenzymes represent “biochemical fossils” from very primitive stages in the history of life (Eschenmoser, 1988). Thus PLP itself, is presumably much older than the separation between the three kingdoms of life. In fact, the nonenzymatic synthesis of pyridoxal under prebiotic conditions is considered possible (Austin & Waddell, 1999; Aylward & Bofinger, 2006) while the presence of a 5’ group could hint to an ancestral attachment of the cofactor to RNA molecules (Jadhav & Yarus, 2002).

But if the general arguments about the antiquity of cofactors apply to PLP, how did PLP-dependent enzymes emerge and evolve? The emergence of a PLP-dependent enzymes family might include the following sequence of events. The first step was the reaction of PLP with a lysine residue of the progenitor protein. This reaction was facilitated by noncovalent binding of the cofactor.

A further prerequisite for the development toward an effective protein catalyst was the pre-existence of a rudimentary substrate-binding site adjacent to the PLP-binding site that allowed the transamination with an amino acid substrate and thus the formation of the planar coenzyme-substrate aldimine adduct. Formation of the external aldimine, the starting point for all PLP-dependent reactions, is an absolute necessity for the catalytic efficacy of PLP toward amino acid substrates.

[¥]Exceptions are animals (which are unable to synthesize PLP *de novo*) and proteobacteria of gamma-subgroup such as *E.coli* (which are postulated to have developed a new and more complex biosynthetic route (Fitzpatrick , 2007).

Subsequent optimization of the noncovalent interactions between the protein and the PLP-substrate adduct may be assumed to have led to improved catalytic efficiency and to specialization for reaction and substrate specificity. In the progenote, the subsequent divergent evolution of the protoenzymes was first determined by functional specialization for reaction as well as substrate specificity and then by the phylogenetic divergence, that is, by speciation (**Figure 16**).

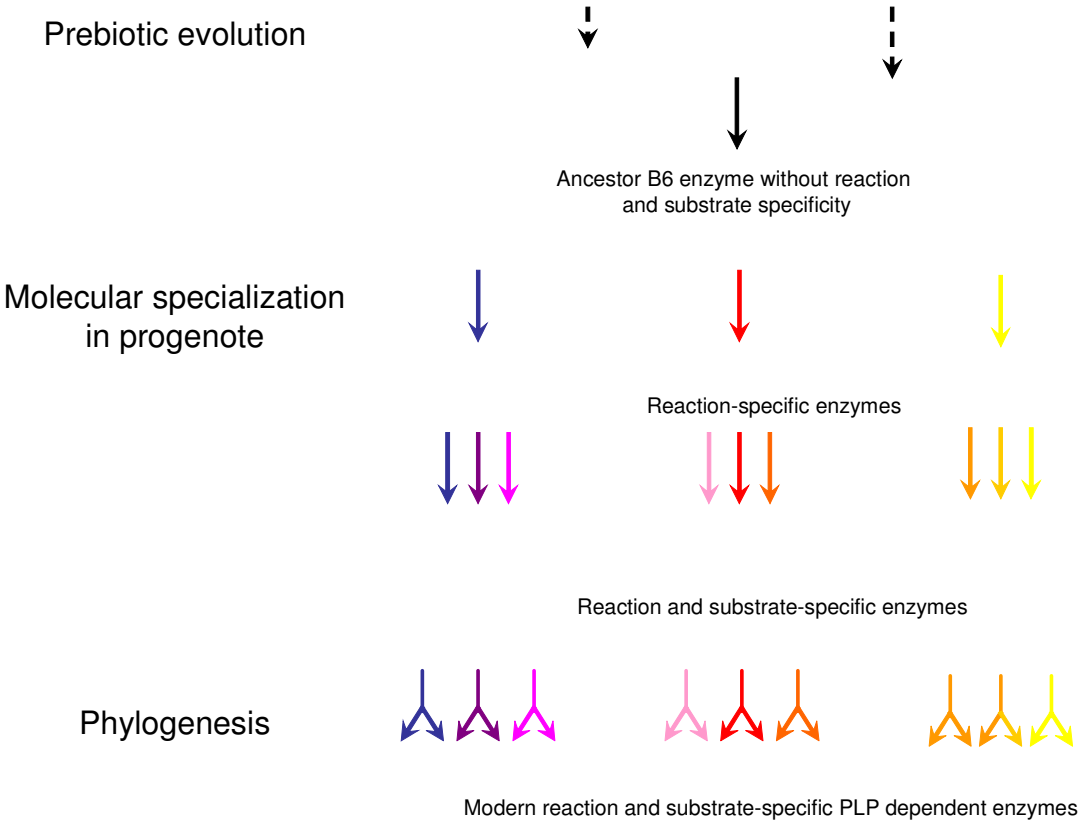


Figure 16. Synopsis of the molecular evolution of PLP-dependent enzymes. The scheme depicts the temporal sequence of the processes that led from the ancestor of the given PLP-dependent enzymes family, which very likely was an allrounder catalyst, to the reaction and substrate-specific enzymes of amino acid metabolism in recent species. The PLP-dependent enzymes are of multiple evolutionary origin, the course of events outlined in the scheme has occurred in parallel a couple of times, each time starting from a different ancestor protein with another fold. For brevity's sake, the development of only one PLP-dependent enzyme family is shown. The main features of molecular evolution are most clearly evident from the data on the large α family (**Figure 17**). However, the data on the other much smaller PLP-dependent enzyme families, (**Figure 18**) (i.e., the β family, as well as the D-alanine aminotransferase and the alanine racemase family) comply with the same scheme. The first cells and last common ancestor of prokaryotes and eukaryotes are estimated to have been in existence about 3500 and 1500-2000 million years ago respectively (Doolittle ,1996).

The process of formation of a primordial PLP-dependent enzyme by reaction of PLP with suitable apoprotein has happened more than once and has led to the emergence of several independent evolutionary lineages.

Family pedigrees and course of functional specialization

Because of its many diverse member enzymes, the evolutionary pedigree of the Fold-type I family is most informative for deducing how functional specialization proceeded.

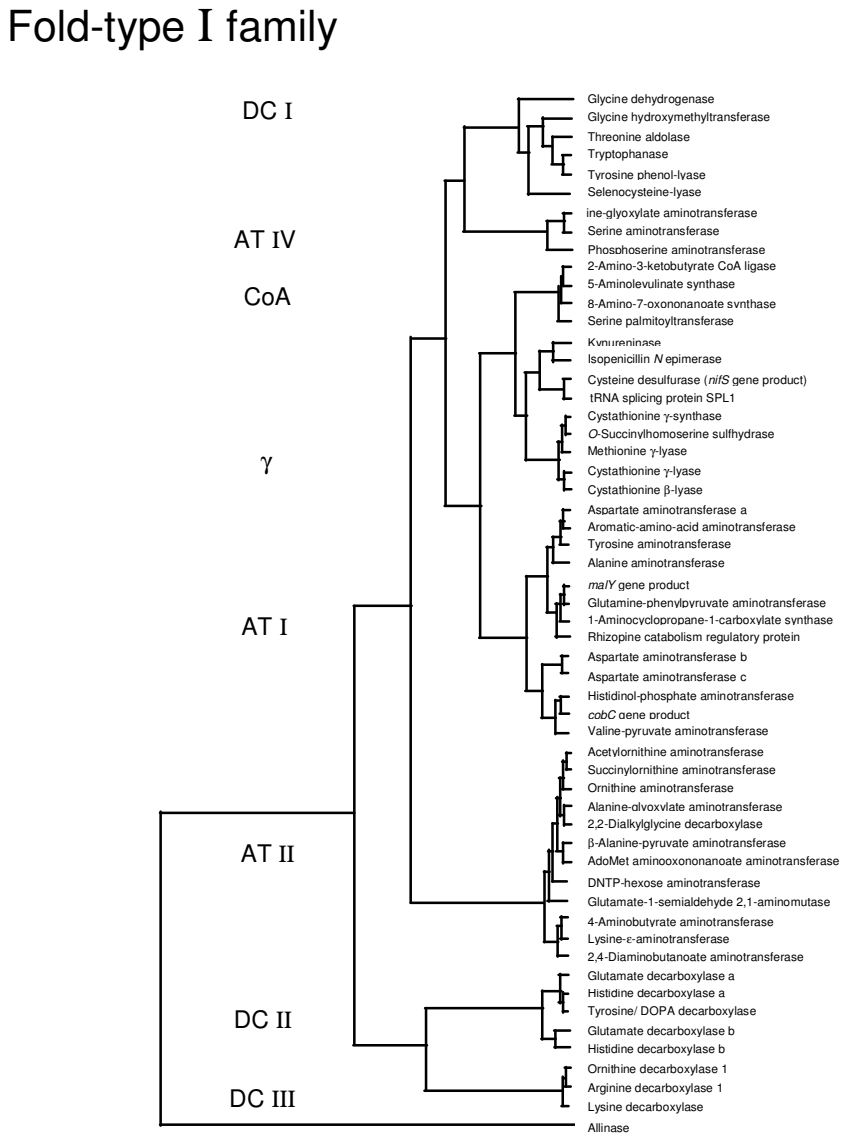
Scrutiny of the pedigree (**Figure 17**) reveals a clear pattern in the temporal sequence of events that led to the functional specialization of these enzymes. Apparently, the ancestor enzyme of the α family was specific for covalency changes limited to Cα.

The region-specific ancestor enzyme then diverged into reaction-specific enzymes, such as the three subfamilies of aminotransferase, the three subfamilies of amino acid decarboxylases, the CoA-dependent acyltransferases like 5-aminolevulinate synthase, etc.

The subdivision of the enzymes into subfamilies of aminotransferases (subfamilies AT I, II, and IV in the Fold-type I family and subfamily AT III in the D-alanine aminotransferase family; Metha,1993), subfamilies of amino acid decarboxylases (subfamilies DC I-III in the Fold-type I family and subfamily DC IV in the alanine racemase family; Sandmeier, 1994), a γ subfamily (Alexander, 1994) and a CoA subfamily (Metha & Christen, 2000). The enzymes that catalyze reactions extending to Cβ and Cγ, for example, the glycine (serine) hydroxymethyltransferase/tyrosine phenol-lyase group or the γ subfamily seem to have diverged from sublineages of the α family just prior to the specialization for substrate specificity. The philogenetic tree shows clearly that the last and also shortest phase was the specialization for substrate specificity.

For some enzymes, a further subdivision of subfamilies into different groups seems appropriate on the basis of their relative degree of relationship among themselves and with other enzymes. The subdivision of subfamily AT I into groups *a-c* of aspartate aminotransferase and of subfamily DC II into groups *a* and *b* of glutamate decarboxylate and groups *a* and *b* of histidine decarboxylase seems by and large to reflect the course of phylogenesis (**Table 1**).

Conceivably, these enzymes were already fully specialized for their particular reaction and their substrate very early in phylogenetic evolution. The 1-Aminocyclopropane-1-carboxylate synthase is as yet the only enzyme in the three AT subfamily of the α family that is not an aminotransferase but rather catalyzes an α,γ-elimination reaction.



Mehta & Christen, Adv Enzymol Relat Areas Mol Biol 2000; 74:129-184

Figure 17. Evolutionary pedigree of the α family. The evolutionary tree was constructed with the GrowTree program on the basis of FPA (for details on the constructions of the evolutionary trees see, Metha and Christen, 2000). The subfamilies are numbered as previously (Mehta, 1993; Sandmeier, 1994; Alexander, 1994). Within subfamilies, some enzymes have been subdivided into different groups, that is, aspartate aminotransferase groups *a* to *c* in subfamily AT I, glutamate decarboxylase groups *a* and *b*, as well as histidine decarboxylase groups *a* and *b* in decarboxylase subfamily DC II, because the sequences of the different groups could not be aligned with each other due to a distant relationship. A different situation is found with ornithine and arginine decarboxylases. The members of ornithine decarboxylase group I have an α family fold and are all of prokaryotic origin. The single known amino acid sequence of arginine decarboxylase group I is from *Escherichia coli* (biodegradative enzyme). However, other forms of these enzymes belong to the alanine racemase family with a $(\alpha/\beta)_8$-barrel fold (ornithine decarboxylase 2 and arginine decarboxylase 2, **Figure 18**).

The other enzymes of the fold-type I (**Figure 17**) that do not catalyze an α reaction, include the γ-family (with one member enzyme, cystathionine β-lyase, catalyzing a β reaction), an evolutionary branch with glycine hydroxymethyltransferase, threonine aldolase, tryptophanase, tyrosine phenol-lyase, selenocysteine synthase, an evolutionary branch with kynureninase, isopenicillin N epimerase, cysteine desulfurase, as well as the very distantly related alliinase.

The pedigrees of the β family, the D-alanine aminotransferase family, and the alanine racemase family are small and do not allow general conclusions (**Figure 18**). The two enzymes of the β family that act on enantiomeric substrates, that is, D- and L-serine dehydratase fall into two different branches of the tree.

Table 1

Species representation in the different groups of subfamily AT I of Aspartate Aminotransferases and subfamily DC II of Glutamate Decarboxylases and Histidine Decarboxylases

		Occurrence in		
Enzyme	Group	Eukaryotes	Eubacteria	Archeabacteria
Aspartate aminotransferase in subfamily AT I	*a*	+	+	
	b		+	+
	c			+
Glutamate decarboxylase in subfamily DC II	*a*	Animals		
	b	Plants	+	+
Histidine decarboxylase in subfamily DC II	*a*	Animals		
	b	Plants	+	

Table 1. All enzymes belong to the α family (**Figure 15**). The subdivision of aspartate aminotransferases subfamily AT I into group *a* and *b* was first proposed by Okamoto (1996) and Nakai (1999).

What were the properties of the hypothetical intermediates of the evolutionary process delineated by the pedigrees of the PLP-dependent enzymes?

The primordial enzyme with covalently bound PLP probably collected L-amino acids of any type and facilitated the formation of PLP-amino acid aldimine adducts thus accelerating all types of PLP-dependent transformations of amino acids.

In a later evolutionary phase, the reaction-specific ancestor enzymes must have favored one particular reaction over all others without differentiating between the various amino acids.

Why came specialization for reaction specificity before specialization for substrate specificity? Two explanations may be offered, one mechanistic and one metabolic.

Mechanistically, specialization of the catalytic apparatus for reaction specificity may be assumed to require more extensive structural adaptations than the modifications of the substrate-binding site necessary to make it interact with a specific substrate. This notion is supported by the relatively short evolutionary time that was used for developing substrate specificity (**Figure 17**).

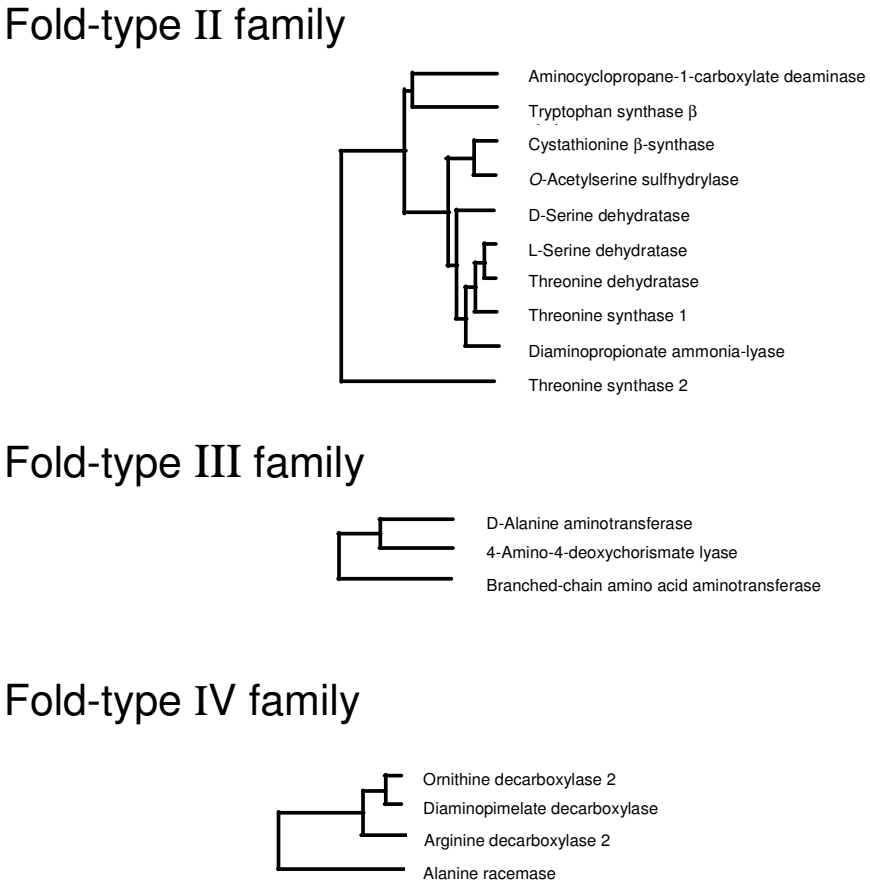
A substrate-unspecific enzyme is not necessarily equivalent to an enzyme with low affinity for the substrates. The relative facility to interact with different substrates is illustrated by the relatively broad substrate specificity of many PLP-dependent enzymes, in particular by the aminotransferases, many of which accept a particular amino acid/oxo acid substrate pair, for example, aromatic or basic, in one half reaction and act on glutamate and 2-oxoglutarate in the second half reaction with reverse direction.

The recommended names of the aminotransferases (**Figure 17**, **18**) are derived from the amino acid substrates for which the given enzyme is considered to be specific. However, it should be borne in mind that for many enzymes the second substrate pair, glutamate and 2-oxoglutarate, provides the link between the metabolism of the amino acid carbon skeleton and nitrogen metabolism.

For the organization of metabolism in the uncompartmented progenote cell, the development of catalysts that accelerate one particular reaction of diverse substrates seems more important than the development of catalyst that act only on one substrate but do not accelerate any of the diverse reactions to an extent exceeding that reached already by the primordial ancestor PLP-dependent enzyme.

Evolutionary precedence of reaction specificity over substrate specificity is also observed in numerous other instances, for example, in families of proteinases or sugar kinases (Perona and Craik, 1997). As corollary of the course of functional specialization, it should be easier to change experimentally the substrate than the reaction specificity of a PLP-dependent enzyme.

This observation is consonant with the conclusion drawn from experiments of proteins engineering (Hedstrom, 1994) that it might be easier first to select an enzyme with the desired catalytic activity and then to tailor its substrate binding site to new substrate instead of trying to adapt pre-existing binding sites for catalysis of a new reaction.



Mehta and Christen , Adv Enzymol Relat Areas Mol Biol 2000; 74: 129-184

Figure 18. Evolutionary pedigrees of the Fold-type II family, Fold-type III family, and Fold-type IV family. Within a given family, the length of the branches is a measure of the relative evolutionary time. In the Fold-type II family, the sequences of threonine synthase fall into two groups (1 and 2). Threonine synthase 1 includes gram-positive bacterial and archebacterial sequences. Threonine synthase 2 includes sequences from mainly gram-negative bacteria and sequences from yeast. This group proved still to be closely related to threonine synthase 1 and to some other members of the Fold-type II family, for example, L-serine dehydratase and cystathionine β-synthase. The Fold-type IV family includes branched-chain amino acid aminotransferase acting on L-valine and L-isoleucine. It is to be noted that this enzyme (EC 2.6.1.42) is distinct from the valine-pyruvate aminotransferase (EC 2.6.1.66) in the α family, the two enzymes use glutamate/2-oxoglutarate and alanine/pyruvate, respectively, as the second substrate pair in the transamination reaction. The Fold-type III family also includes three decarboxylases acting with basic amino acids substrates. Ornithine decarboxylase 2 (OrnDC2), which appears to be exclusively eukaryotes, and arginine decarboxylases 2 (ArgDC2), which has been found to occur in both eukaryotic and eubacteria, represent the second group of the respective enzymes. Groups OrnDC1 and ArgDC1 are part of the Fold-type I family (**Figure 17**).

Divergence in catalytic mechanisms

In view of the numerous constraints apparently due to the particular chemical properties of PLP and the mechanistic similarity of its interactions with the amino acid substrates, it seems remarkable that the following mechanistic features vary rather widely (Jansonius, 1998):

1. The noncovalent interactions between the phosphate group of the coenzyme and the apoprotein are quite different both in type and number, resulting in different degrees of local electrostatic compensation of the negative charge of the phosphate group.

2. The group interacting with *N1* of the pyridine ring is the distal carboxylate group of an aspartate or glutamate residue in all known enzyme structures of the Fold-type I family, the Fold-type IV family and in orntihine decarboxylase of Fold-type III family; in the Fold-type II enzymes and in alanine racemase, however, it is a serine and an arginine residue, respectively. Apparently, the stabilization of a positive charge at *N1* is not a prerequisite for PLP catalysis. The effect of the group interacting with *N1* on the electron distribution within the cofactor has been experimentally probed by site-directed mutagenesis experiments with aspartate aminotransferase (Yano, 1992; Onuffer & Kirsch, 1994) and the β-subunit of tryptophan synthase (Jhee, 1998).

3. K^+ ions have been found to be essential for catalytic activity of particular PLP-dependent enzyme such as 2,2-dialkylglycine decarboxylase (Toney,1993, 1995), tryptophan synthase β (Peracchi,1995), tyrosine phenol-lyase (Sundararaju, 1997), and tryptophan phenol-lyase (Isupov, 1998). In most PLP-dependent enzymes, however, monovalent cations play neither a mechanistic nor structural role.

4. Virtually all PLP-dependent enzymes seem to be oligomeric proteins, the degree of oligomerization ranging from dimers such as aspartate aminotransferases to dodecameric ornithine decarboxylase of prokaryotes (Momany, 1995). The only PLP-dependent enzymes that has been reported to be a monomer is D-serine dehydratase (Dowhan & Snell, 1970) and also threonine synthase (Omi, 2003).

On the basis of the present knowledge on the structure-function relationship of PLP-dependent enzymes, it does not seem possible to attribute a functional significance to the above divergent features. They might have arisen by chance in the different evolutionary lineages, in contrast to the features that are common to all PLP-dependent enzymes.

Time of emergence and rate of evolution

In three B6 enzyme families, enzymes are found of which sequences from all biological kingdoms are available (**Table 2**). For each enzyme, its sequences in archebacteria, eubacteria, and eukaryotes are all homologous with each other and are more closely related among each other than with the sequences of any other PLP-dependent enzymes in the three biological kingdoms. This suggest that these enzymes were already completely developed, that is, endowed with reaction and substrate specificity, in the universal ancestor cell.

Table 2

Family	Enzyme (EC Number)
α	Glycine hydroxymethyltransferase (2.1.2.1)
	8-Amino-7-oxononanoate synthase (2.3.1.47)
	Histidinol-phosphate aminotransferase (2.6.1.9)
	Acetylornithine aminotransferase (2.6.1.11)
	Phosphoserine aminotransferase (2.6.1.52)
	AdoMet 8-amino-7-oxonanoate aminotransferase (2.6.1.62)
	1-Aminocyclopropane-1-carboxylate deaminase (4.1.99.4)
β	Tryptophan synthase β chain (4.2.1.20)
	Threonine synthase (4.2.99.2)
D-alanine aminotransferase	Branched-chain amino acid aminotransferase (2.6.1.42)

Table 2. Enzymes with Homologous Sequences in all three biological kingdom.

These enzymes must have arisen and specialized before the three biological kingdoms branched off from each other 1500 to 1000 million years ago. An emergence so early in biological evolution probably applies to numerous other PLP-dependent enzymes, e.g. other aminotransferases and amino acid synthases, as they catalyze reactions that link different areas of metabolism and must have been essential in organizing the metabolic pathways in the progenote. This notion of an early organization of amino acid metabolism is consonant with a general study on the phylogenetic distribution of protein families into the three biological kingdoms. The analysis showed that in the last universal ancestor of contemporary cells the major metabolic pathways were already established (Ouzounis & Kyrpides, 1996).

Functional specialization during phylogenesis

In some instances, however, specialization for (or change of) substrate specificity seems to have overlapped with phylogenetic evolution. Within the aminotransferase subfamily AT I, the aspartate aminotransferases sequences are subdivided into three groups *a-c* (**Figure 16**). Aspartate aminotransferase group *a* includes eukaryotic sequences, that is of higher vertebrates and plants, and certain eubacterial sequences. Aspartate aminotransferase group *b* consists exclusively of prokaryotic sequences, some of them of hyperthermophilic and archebacterial species. Aspartate aminotransferase group *c* is exclusively archebacterial (**Table 1**).

The sequences of group *a* prove to be more closely related to those of aromatic-amino-acid aminotransferase, tyrosine aminotransferase, and alanine aminotransferase in the same evolutionary branch than to the aspartate aminotransferases in group b and c. The latter groups, though distantly related with group a, clearly constitute evolutionary branches on their own.

Apparently, in the case of these aminotransferase the specialization for substrates specificity occurred in parallel with early phases of phylogenetic evolution when the three kingdoms diverged.

Another example of molecular evolution concomitant with phylogenesis and perhaps due to functional constraints specific for a particular phylogenetic branch, is also observed in the amino acid decarboxylase subfamily DC II.

Glutamate decarboxylase group *a* is more closely related to histidine decarboxylase group *a* than to glutamate decarboxylase group *b*. Glutamate decarboxylase group *a* occurs exclusively in animals; apparently it developed after the divergence of animals from plants and lower eukaryotes and might perhaps relate to the newly acquired function of the decarboxylation product 4-aminobutyrate (GABA) as a neurotransmitter. A third example of specialization during phylogenetic evolution are the two threonine synthase groups in the β family (**Figure 17**). How fast did PLP-dependent enzymes, after their specialization for reaction and substrate specificity, change their amino acid sequence? Two prerequisites have to be met in order to explore this question: enough sequences that can be reliably aligned and a reliable time scale of phylogenetic divergence. Screening of the PLP-dependent enzymes sequence database with these two criteria allowed estimation of the rate of evolution of a number of PLP-dependent enzymes.

The values cover a relatively broad range from the fast-changing serine pyruvate aminotransferase with a unit evolutionary period (or UEP)[¥] of only 4.6 million years, which is slightly longer than the 3.7 million years for the α chain of haemoglobin, to the slowly changing glutamate decarboxylase with 45 million years, which compares with other relatively slowly changing intracellular enzymes such as triosephosphate isomerase and glutamate dehydrogenase with unit evolutionary periods of 20 and 55 million years, respectively (Wilson, 1977; Salzmann, 2000). The glutamate decarboxylase sequences that were considered are all from mammalian species in which this enzyme plays a delicate role as producer of GABA (the most important inhibitory transmitter in the central nervous system). Glutamate decarboxylase is found only in neurons that use 4-aminobutyrate as transmitter.

Rate of evolution and subgroup affiliation do not seem to correlate. Fast-changing serine pyruvate aminotransferase and slowly changing glutamate decarboxylase belong both to the α family and the evolutionary rates of the enzymes of the other families seem to intermingle at random with those of the α enzymes. The differential rates of neutral evolution of the PLP-dependent enzymes do not correlate with their different folds. Apparently, the rates of evolution are largely determined by functional constraints.

[¥]UEP: defined as the number of years required to cause a 1% change in evolutioarily divergen sequences.

Genomic distribution of PLP-dependent enzymes

A recent study (**Figure 19A**), on PLP-dependent enzymes in complete genomes, demonstrated that nearly 1.5% of all genes in most prokaryotic genomes encode PLP-dependent enzymes and that 4% of the 140 distinct enzymes classified by the Enzyme Commission depend on PLP (Percudani & Peracchi, 2003).

The number of genes for PLP-dependent enzymes in extant, free-living microorganis presumably depends on their adaptation to specific nutrient sources. Only two EC-classified activities are present in all the available genomes of free-living organisms: aspartate aminotransferase (EC 2.6.1.1) and serine hydroxymethyltransferase (EC 2.1.2.1).

If the importance of PLP as a biological cofactor in prokaryotes is underscored by the high fraction of genes encoding PLP-dependent enzymes, this fraction decrease with expansion of the size and complexity of the genome (**Figure 19B**), perhaps because PLP-dependent enzymes are mostly involved in basic metabolic pathways rather than in more specialized regulatory functions.

Although the absolute number of PLP-dependent enzymes is somewhat larger in higher eukaryotes than in microorganisms (**Figure 19A**), the number of EC-classified activities that are represented is not (**Figure 19C**). This can be explained in part by the occurrence of organelle-specific or tissue-specific isozymes, which are encoded by different genes but have the same enzymatic activity.

However, the relatively modest number of classified activities that are found in higher eukaryotes might also imply that some gene products assigned to the same EC number on the basis of sequence homology actually have different activities.

Limitations to the identification of PLP enzymes

The current picture of the genomic distribution of PLP-dependent enzymes is informative, but it is also preliminary, and to some degree incomplete. This is partly due to limits that are inherent to homology searches, which may fail to recognize PLP-dependent genes in at least two situations.

First, it is possible that structurally similar enzymes escape detection if their sequence similarity has become negligible. A case in point is provided by the crystal structure of the yeast hypothetical protein YBL036C.

This molecule was not predicted to be a PLP-dependent enzyme and was even suggested to possess a novel protein fold. However, the solution of the crystal structure

revealed a classic $(\alpha/\beta)_8$ barrel, with pyridoxal phosphate covalently bound at the bottom. Thus, YBL036C closely resembles PLP-dependent enzymes such as alanine racemase, and it has indeed been shown to possess some amino-acid racemase activity (Eswaramoorthy, 2003).

A second case in which homology searches can fail to identify PLP-dependent genes is when the enzymes that they encode do not fall into the five fold-type categories described above and do not resemble any of the known PLP-dependent enzymes. For example, some recently sequenced bacterial aminomutase (such as D-lysine-5,6-aminomutase) which use vitamin B12 cofactor in addition to PLP, show no significant similarity to other PLP-dependent enzymes at the sequence level, but instead resemble some B12-containing proteins (Chang & Frey, 2000; Chen, 2001).

Limitations to the functional assignment of PLP enzymes

A substantial fraction of the genes that have been identified as encoding PLP-dependent enzymes remain functionally unclassified or only tentatively classified; for example, there is no assignment of catalytic activity for about 20% of the putative PLP-dependent enzymes encoded by the human genome, and one-third of all PLP-dependent enzyme classified by the Enzyme Commission are still uncharacterized in terms of sequence (Metha & Christen, 2000).

These number are enough to evidence the lack of information regarding the family of PLP-dependent enzymes, and, on the other side, how much work is needed before to fully exploit the wealth of information contained in the genome sequences.

In fact this number might even be an underestimate because, although bioinformatics tools are well suited to detecting structural similarities and determining phylogenetic relationships, they are much less reliable for establishing protein function solely based on homology, even when sequence identity is > 50% (Thornton, 2000; Rost, 2002).

This problem is particularly severe for PLP-dependent enzymes, as their catalytic mechanisms almost invariably involve the formation of carbanionic intermediates that are stabilized by the cofactor.

Such constant mechanistic features have presumably favoured the apparence of enzymes with identical or highly similar activities several times during the course of evolution (Mehta & Christen, 2000).

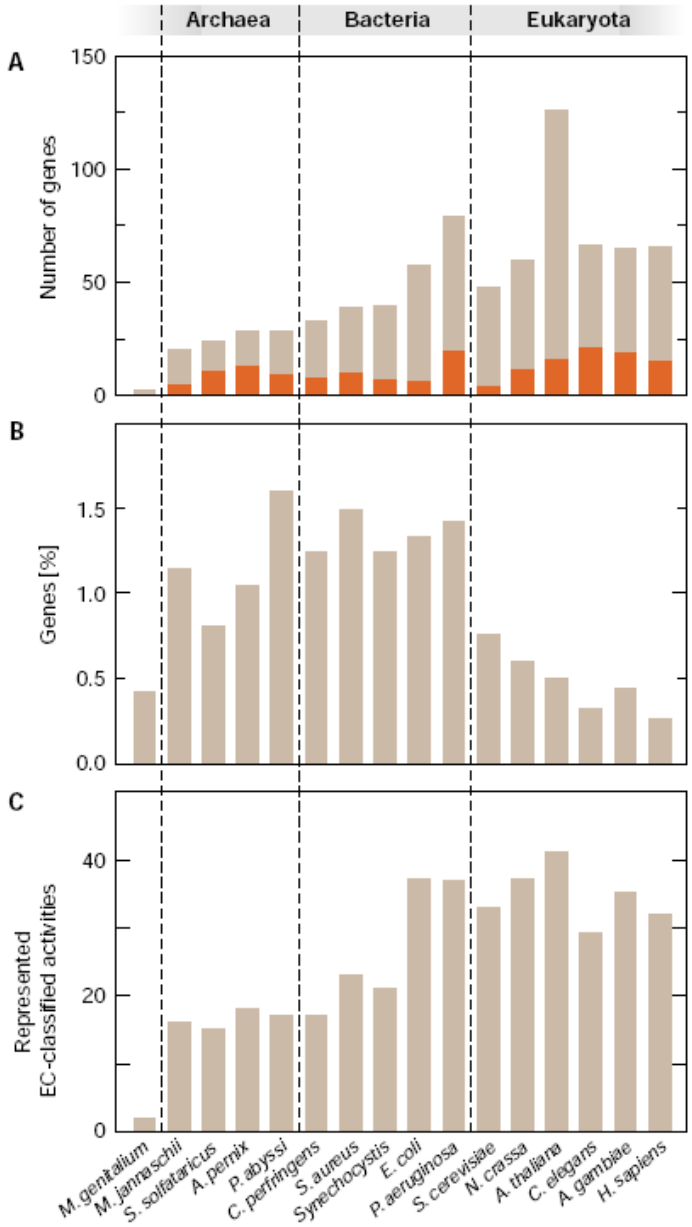


Figure 19. Distribution of genes that encode pyridoxal-phosphate-dependent enzymes in representative genomes. The organisms used as examples are: *Mycoplasma genitalium* (not a free-living organism),*Methanococcus jannaschii, Sulfolobus solfataricus, Aeropyrum pernix, Pyrococcus abyssi, Clostridium perfringens, Staphylococcus aureus, Synechocystis* sp. PCC 6803*, Escherichia coli, Pseudomonas aeruginosa*, *Saccharomyces cerevisiae*, *Neurospora crassa*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Anopheles gambiae* and *Homo sapiens*. (**A**) Total number of pyridoxal-phosphate (PLP)-dependent genes per genome; the fraction of sequences that could not be classified in terms of function, on the basis of our search criteria, is shown in red. The relatively high number of PLP-dependent genes in *A. thaliana* can be explained in part by the occurrence, for many enzymes, of at least three isoforms (cytoplasmic,mitochondrial and plastidic) that are encoded by different genes. (**B**) The percentage of genes in each genome that encode PLP-dependent enzymes. (**C**) The number of Enzyme Commission (EC)-classified PLP-dependent activities for which sequences were found in each genome.Note that similar numbers of activities can be found in organisms with very different assortments of enzymes. For example, only about one-third of the PLP-dependent activities found in *E. coli* are also found in *H. sapiens*, despite the fact that the two organisms have a similar absolute number of activities.

In addition, PLP-dependent enzymes are involved in a surprising variety of cellular processes, so that even for an enzyme for which the catalytic activity can be assigned with high confidence on the basis of its sequence, the actual biological function may remain uncertain. Enzymes that have the same catalytic activity but are evolutionarily unrelated and functionally diverse are exemplified by a group of PLP-dependent enzymes with desulfhydrase activity, that is, enzymes that release H_2S from thiol-containing amino acids. Bacteria use desulfhydrase not only in amino acids metabolism and in adaption to new nutrient sources (Soutourina, 2001), but also sometimes as virulence factors (Krupka, 2000; Fukamachi, 2002). One *Escherichia coli* enzyme with desulfhydrase activity is even known to act as a modular of gene expression, although this function seems to be unrelated to catalysis (Clausen , 2000). Sulfhide production by PLP-dependent enzymes is also important in vertebrates, in which H_2S has been shown to function as a neuromodulator (Kimura, 2002). PLP-dependent enzymes with desulfhydrase activity come from several evolutionary lineages. For example, although most bacterial L-cysteine desulfhydrases belong to the fold-type I group, cystathionine β-synthase (the enzyme responsible for H_2S production in the mammalian brain) is a fold-type II enzyme. A L-cysteine desulfhydrase from *Fusobacterium nucleatum* is also a fold-type II enzyme, and its closest sequence homologue is a cysteine synthase (Fukamachi,2002). D-cysteine desulfhydrase from *E.coli* is not closely related to other desulfhydrase, and is most similar to 1-aminocyclopropane-1-carboxylate deaminase (Soutourina, 2001).

In search of novel PLP-dependent enzyme

Even in the cases of model organism, the difficulties in establishing function on the basis of gene genealogies imply that much experimental work will be required to achieve a genome-wide classification of all PLP-dependent enzymes. With the advancement in functional genomics it is likely to expect the discovery of many ‘novel’ PLP-dependent enzymes that have activities that are not yet described or characterized.

In fact, the recent literature supports this assumption. One example is provided by the identification of *Salmonella typhimurium* gene that encodes an enzyme with L-threonine-*O*-3-phosphate decarboxylase activity (Brushaber, 1998). This was the first enzyme described as having this activity, and its discovery shed new light on the pathway that leads to cobalamin biosynthesis in *Salmonella*. The sequence of this enzyme is only distantly related to those of other PLP-dependent decarboxylase, but is highly similar to those of histidinol-phosphate aminotransferases (Brushaber, 1998).

In another case, (Wolosker, 1999) described a PLP-dependent serine racemase that is expressed in the human brain. The existence of this enzyme had been previously overlooked, as D-amino acids were not known to have any metabolic or physiological function in vertebrates. However, it is now recognized that D-serine is a neurotransmitter (Wolosker, 1999; De Miranda, 2000). The human serine racemase does not show sequence similarity to bacterial amino acid racemases (fold-type III), but instead resembles threonine ammonia-lyases, which are fold-type II enzymes (De Miranda, 2000). The discovery of new enzymes that produce chemical mediators in higher eukaryotes may have interesting implications for the recycling of certain proteins for new physiological functions. Indeed, plant and animal cell often use PLP-dependent enzymes for the synthesis of hormones and chemical messengers, and these enzymes can be close homologues of enzymes involved in basic metabolic pathways.

Orphan PLP-dependent activities and catalytic promiscuity

Some PLP-dependent enzymes are still uncharacterized at the sequence level because they exist only in obscure organisms, or because their function is not of sufficient scientific interest. However, these reasons doesn't explain the large number of 'orphan' PLP-dependent activities.

A probable explanation refer to the phenomenon of 'catalytic promiscuity', that is the ability of a single enzyme to catalyze different chemical reactions (O'Brien & Herschlag, 1999). This phenomenon is particularly frequent among PLP-dependent enzymes. Due to their common mechanistic features the occurrence of catalytic promiscuity, in addition to a low substrate specificity, suggest that an organism may have less genes encoding PLP-dependent enzymes than PLP-dependent activities (Percudani & Peracchi, 2003).

Furthermore, the possibility of alternative splicing in higher eukaryotes, this can increase the functional diversity of proteins in general (Graveley, 2001). It must be stressed, however, that although many alternatively spliced PLP-dependent genes have been described in animal cells, splice variants that differ significantly in terms of reaction or substrate specificity have not been reported.

Moreover, in some cases it has been shown that certain splice variants cannot bind PLP and presumably do not function as enzymes (Bond, 1990; Liu, 2001).

References

Adams, D. O. and S. F. Yang (1979). "Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene." Proc Natl Acad Sci U S A **76**(1): 170-174.

Alexander, F. W., E. Sandmeier, et al. (1994). "Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families." Eur J Biochem **219**(3): 953-60.

Ana Rios, T. L. A., and John P. Richard (2000). "Formation and Stability of Organic Zwitterions in Aqueous Solution: Enolates of the Amino Acid Glycine and Its Derivatives." Journal of America Chemistry Society **122**(39): 9373-85.

Bond, R. W., R. J. Wyborski, et al. (1990). "Developmentally regulated expression of an exon containing a stop codon in the gene for glutamic acid decarboxylase." Proc Natl Acad Sci U S A **87**(22): 8771-5.

Brushaber, K. R., G. A. O'Toole, et al. (1998). "CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (R)-1-amino-2-propanol O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in Salmonella typhimurium LT2." J Biol Chem **273**(5): 2684-91.

Brzovic, P., E. L. Holbrook, et al. (1990). "Reaction mechanism of Escherichia coli cystathionine gamma-synthase: direct evidence for a pyridoxamine derivative of vinylglyoxylate as a key intermediate in pyridoxal phosphate dependent gamma-elimination and gamma-replacement reactions." Biochemistry **29**(2): 442-51.

Burkhard, P., G. S. Rao, et al. (1998). "Three-dimensional structure of O-acetylserine sulfhydrylase from Salmonella typhimurium." J Mol Biol **283**(1): 121-33.

Chang, C. H. and P. A. Frey (2000). "Cloning, sequencing, heterologous expression, purification, and characterization of adenosylcobalamin-dependent D-lysine 5, 6-aminomutase from Clostridium sticklandii." J Biol Chem **275**(1): 106-14.

Chen, H. P., S. H. Wu, et al. (2001). "Cloning, sequencing, heterologous expression, purification, and characterization of adenosylcobalamin-dependent D-ornithine aminomutase from Clostridium sticklandii." J Biol Chem **276**(48): 44744-50.

Christen, P. and P. K. Mehta (2001). "From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Chem Rec **1**(6): 436-47.

Clausen, T., A. Schlegel, et al. (2000). "X-ray structure of MalY from Escherichia coli: a pyridoxal 5'-phosphate-dependent enzyme acting as a modulator in mal gene expression." Embo J **19**(5): 831-42.

Corey, E. S., RA (1956). "Stereoelectronic Control in Enolization-Ketonization Reactions." Journal of America Chemistry Society **78**(24): 6269-6278.

Cronin, C. N. and J. F. Kirsch (1988). "Role of arginine-292 in the substrate specificity of aspartate aminotransferase as examined by site-directed mutagenesis." Biochemistry **27**(12): 4572-9.

De Miranda, J., A. Santoro, et al. (2000). "Human serine racemase: moleular cloning, genomic organization and functional analysis." Gene **256**(1-2): 183-8.

Denessiouk, K. A., A. I. Denesyuk, et al. (1999). "Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes." Proteins **35**(2): 250-61.

Doolittle, R. F., D. F. Feng, et al. (1996). "Determining divergence times of the major kingdoms of living organisms with a protein clock." Science **271**(5248): 470-7.

Dowhan, W., Jr. and E. E. Snell (1970). "D-serine dehydratase from Escherichia coli. II. Analytical studies and subunit structure." J Biol Chem **245**(18): 4618-28.

Dunathan, H. C. (1966). "Conformation and reaction specificity in pyridoxal phosphate enzymes." Proc Natl Acad Sci U S A **55**(4): 712-6.

Dunathan, H. C. and J. G. Voet (1974). "Stereochemical evidence for the evolution of pyridoxal-phosphate enzymes of various function from a common ancestor." Proc Natl Acad Sci U S A **71**(10): 3888-91.

Eswaramoorty, S. (1999). "Crystal structure of a yeast hypotetical protein."

Eswaramoorty, S., Gerchman, S., Graziano, V., Kycia, H., Studier, F.W., Swaminathan, S. (2003). "Structure of a yeast hypothetical protein selected by a structural genomics approach." *Acta Crystallograph.D Biol. Crystallogr.* **59**: 127-135.

Fitch, W. M. (1970). "Distinguishing homologous from analogous proteins." Syst Zool **19**(2): 99-113.

Fukamachi, H., Y. Nakano, et al. (2002). "Cloning and characterization of the L-cysteine desulfhydrase gene of Fusobacterium nucleatum." FEMS Microbiol Lett **215**(1): 75-80.

Gallagher, D. T., G. L. Gilliland, et al. (1998). "Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase." Structure **6**(4): 465-75.

Goldberg, J. M. and J. F. Kirsch (1996). "The reaction catalyzed by Escherichia coli aspartate aminotransferase has multiple partially rate-determining steps, while that catalyzed by the Y225F mutant is dominated by ketimine hydrolysis." Biochemistry **35**(16): 5280-91.

Graveley, B. R. (2001). "Alternative splicing: increasing diversity in the proteomic world." Trends Genet **17**(2): 100-7.

Grishin, N. V., M. A. Phillips, et al. (1995). "Modeling of the spatial structure of eukaryotic ornithine decarboxylases." Protein Sci **4**(7): 1291-304.

Han, Q., J. Fang, et al. (2001). "Kynurenine aminotransferase and glutamine transaminase K of Escherichia coli: identity with aspartate aminotransferase." Biochem J **360**(Pt 3): 617-23.

Haruyama, K., T. Nakai, et al. (2001). "Structures of Escherichia coli histidinol-phosphate aminotransferase and its complexes with histidinol-phosphate and N-(5'-phosphopyridoxyl)-L-glutamate: double substrate recognition of the enzyme." Biochemistry **40**(15): 4633-44.

Hayashi, H. (1995). "Pyridoxal enzymes: mechanistic diversity and uniformity." J Biochem (Tokyo) **118**(3): 463-73.

Hayashi, H., K. Inoue, et al. (1993). "Escherichia coli aromatic amino acid aminotransferase: characterization and comparison with aspartate aminotransferase." Biochemistry **32**(45): 12229-39.

Hedstrom, L. (1994). "Engineering for redesign." Current Opinion in Structural Biology **4**(4): 608-611.

Herschlag, D. (1988). " The role of induced fit and conformational changes of enzymes in specificity and catalysis." Bioorganic Chemistry **16**: 62-96.

Heyl, D. L., E; Harris, SA; Folkers, K (1951). "Phosphates of the Vitamin B6 Group I. The Structure of Codecarboxylase." Journal of America Chemistry Society **73**: 3430-33.

Hill, R. E., K. Himmeldirk, et al. (1996). "The biogenetic anatomy of vitamin B6. A 13C NMR investigation of the biosynthesis of pyridoxol in Escherichia coli." J Biol Chem **271**(48): 30426-35.

Hyde, C. C., S. A. Ahmed, et al. (1988). "Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium." J Biol Chem **263**(33): 17857-71.

Isupov, M. N., A. A. Antson, et al. (1998). "Crystal structure of tryptophanase." J Mol Biol **276**(3): 603-23.

Jansonius, J. N. (1998). "Structure, evolution and action of vitamin B6-dependent enzymes." Curr Opin Struct Biol **8**(6): 759-69.

Jeffery, C. J. (1999). "Moonlighting proteins." Trends Biochem Sci **24**(1): 8-11.

Jensen, R. A. and W. Gu (1996). "Evolutionary recruitment of biochemically specialized subdivisions of Family I within the protein superfamily of aminotransferases." J Bacteriol **178**(8): 2161-71.

Jhee, K. H., L. H. Yang, et al. (1998). "Mutation of an active site residue of tryptophan synthase (beta-serine 377) alters cofactor chemistry." J Biol Chem **273**(19): 11417-22.

John, R. A. (1995). "Pyridoxal phosphate-dependent enzymes." Biochim Biophys Acta **1248**(2): 81-96.

Kack, H., J. Sandmark, et al. (1999). "Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes." J Mol Biol **291**(4): 857-76.

Kern, A. D., M. A. Oliveira, et al. (1999). "Structure of mammalian ornithine decarboxylase at 1.6 A resolution: stereochemical implications of PLP-dependent amino acid decarboxylases." Structure **7**(5): 567-81.

Kimura, H. (2002). "Hydrogen sulfide as a neuromodulator." Mol Neurobiol **26**(1): 13-9.

Kirsch, J. F., G. Eichele, et al. (1984). "Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure." J Mol Biol **174**(3): 497-525.

Koshland, D. E. (1958). "Application of a Theory of Enzyme Specificity to Protein Synthesis." Proc Natl Acad Sci U S A **44**(2): 98-104.

Krupka, H. I., R. Huber, et al. (2000). "Crystal structure of cystalysin from Treponema denticola: a pyridoxal 5'-phosphate-dependent protein acting as a haemolytic enzyme." Embo J **19**(13): 3168-78.

Laber, B., K. P. Gerbling, et al. (1994). "Mechanisms of interaction of Escherichia coli threonine synthase with substrates and inhibitors." Biochemistry **33**(11): 3413-23.

Lam, H. M. and M. E. Winkler (1990). "Metabolic relationships between pyridoxine (vitamin B6) and serine biosynthesis in Escherichia coli K-12." J Bacteriol **172**(11): 6518-28.

Liu, X., D. M. Szebenyi, et al. (2001). "Lack of catalytic activity of a murine mRNA cytoplasmic serine hydroxymethyltransferase splice variant: evidence against alternative splicing as a regulatory mechanism." Biochemistry **40**(16): 4932-9.

Malashkevich, V. N., J. J. Onuffer, et al. (1995). "Alternating arginine-modulated substrate specificity in an engineered tyrosine aminotransferase." Nat Struct Biol **2**(7): 548-53.

Matsui, I., E. Matsui, et al. (2000). "The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from Pyrococcus horikoshii." J Biol Chem **275**(7): 4871-9.

Mehta, P. K. and P. Christen (2000). "The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Adv Enzymol Relat Areas Mol Biol **74**: 129-84.

Mehta, P. K., T. I. Hale, et al. (1993). "Aminotransferases: demonstration of homology and division into evolutionary subgroups." Eur J Biochem **214**(2): 549-61.

Metzler, C. M., Cahill Allen, Metzler E.C. (1980). "Equilibria and Absorption Spectra of Schiff Bases." J. Am. Chem. Soc. **102**: 6075-6082.

Metzler, C. M., A. G. Harris, et al. (1988). "Spectroscopic studies of quinonoid species from pyridoxal 5'-phosphate." Biochemistry **27**(13): 4923-33.

Miles, E. W. (2001). "Tryptophan synthase: a multienzyme complex with an intramolecular tunnel." Chem Rec **1**(2): 140-51.

Momany, C., S. Ernst, et al. (1995). "Crystallographic structure of a PLP-dependent ornithine decarboxylase from Lactobacillus 30a to 3.0 A resolution." J Mol Biol **252**(5): 643-55.

Nakai, T., K. Okada, et al. (1999). "Structure of Thermus thermophilus HB8 aspartate aminotransferase and its complex with maleate." Biochemistry **38**(8): 2413-24.

O'Brien, P. J. and D. Herschlag (1999). "Catalytic promiscuity and the evolution of new enzymatic activities." Chem Biol **6**(4): R91-R105.

Okada, K., K. Hirotsu, et al. (1997). "Three-dimensional structure of Escherichia coli branched-chain amino acid aminotransferase at 2.5 A resolution." J Biochem (Tokyo) **121**(4): 637-41.

Okamoto, A., S. Ishii, et al. (1999). "The active site of Paracoccus denitrificans aromatic amino acid aminotransferase has contrary properties: flexibility and rigidity." Biochemistry **38**(4): 1176-84.

Okamoto, A., R. Kato, et al. (1996). "An aspartate aminotransferase from an extremely thermophilic bacterium, Thermus thermophilus HB8." J Biochem (Tokyo) **119**(1): 135-44.

Okamoto, A., Y. Nakai, et al. (1998). "Crystal structures of Paracoccus denitrificans aromatic amino acid aminotransferase: a substrate recognition site constructed by rearrangement of hydrogen bond network." J Mol Biol **280**(3): 443-61.

Onuffer, J. J. and J. F. Kirsch (1994). "Characterization of the apparent negative co-operativity induced in Escherichia coli aspartate aminotransferase by the replacement of Asp222 with alanine. Evidence for an extremely slow conformational change." Protein Eng **7**(3): 413-24.

Onuffer, J. J. and J. F. Kirsch (1995). "Redesign of the substrate specificity of Escherichia coli aspartate aminotransferase to that of Escherichia coli tyrosine aminotransferase by homology modeling and site-directed mutagenesis." Protein Sci **4**(9): 1750-7.

Onuffer, J. J. and J. F. Kirsch (1995). "Redesign of the substrate specificity of Escherichia coli aspartate aminotransferase to that of Escherichia coli tyrosine aminotransferase by homology modeling and site-directed mutagenesis." Protein Sci **4**(9): 1750-7.

Oue, S., A. Okamoto, et al. (1999). "Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues." J Biol Chem **274**(4): 2344-9.

Oue, S., A. Okamoto, et al. (2000). "Cocrystallization of a mutant aspartate aminotransferase with a C5-dicarboxylic substrate analog: structural comparison with the enzyme-C4-dicarboxylic analog complex." J Biochem (Tokyo) **127**(2): 337-43.

Ouzounis, C. and N. Kyrpides (1996). "The emergence of major cellular processes in evolution." FEBS Lett **390**(2): 119-23.

Palm, D., H. W. Klein, et al. (1990). "The role of pyridoxal 5'-phosphate in glycogen phosphorylase catalysis." Biochemistry **29**(5): 1099-107.

Pasternak, A., A. White, et al. (2001). "The energetic cost of induced fit catalysis: Crystal structures of trypsinogen mutants with enhanced activity and inhibitor affinity." Protein Sci **10**(7): 1331-42.

Percudani, R. and A. Peracchi (2003). "A genomic overview of pyridoxal-phosphate-dependent enzymes." EMBO Rep **4**(9): 850-4.

Perona, J. J. and C. S. Craik (1997). "Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold." J Biol Chem **272**(48): 29987-90.

Post, C. B. and W. J. Ray, Jr. (1995). "Reexamination of induced fit as a determinant of substrate specificity in enzymatic reactions." Biochemistry **34**(49): 15881-5.

Rao, N. A., R. Talwar, et al. (2000). "Molecular organization, catalytic mechanism and function of serine hydroxymethyltransferase--a potential target for cancer chemotherapy." Int J Biochem Cell Biol **32**(4): 405-16.

Rhee, S., K. D. Parris, et al. (1996). "Exchange of K+ or Cs+ for Na+ induces local and long-range changes in the three-dimensional structure of the tryptophan synthase alpha2beta2 complex." Biochemistry **35**(13): 4211-21.

Rothman, S. C. and J. F. Kirsch (2003). "How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase." J Mol Biol **327**(3): 593-608.

Salzmann, D., P. Christen, et al. (2000). "Rates of evolution of pyridoxal-5'-phosphate-dependent enzymes." Biochem Biophys Res Commun **270**(2): 576-80.

Sandmeier, E., T. I. Hale, et al. (1994). "Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases." Eur J Biochem **221**(3): 997-1002.

Schirch, V., K. Shostak, et al. (1991). "The origin of reaction specificity in serine hydroxymethyltransferase." J Biol Chem **266**(2): 759-64.

Schneider, G., H. Kack, et al. (2000). "The manifold of vitamin B6 dependent enzymes." Structure **8**(1): R1-6.

Shaw, J. P., G. A. Petsko, et al. (1997). "Determination of the structure of alanine racemase from Bacillus stearothermophilus at 1.9-A resolution." Biochemistry **36**(6): 1329-42.

Soutourina, J., S. Blanquet, et al. (2001). "Role of D-cysteine desulfhydrase in the adaptation of Escherichia coli to D-cysteine." J Biol Chem **276**(44): 40864-72.

Sprang, S. and R. J. Fletterick (1979). "The structure of glycogen phosphorylase alpha at 2.5 A resolution." J Mol Biol **131**(3): 523-51.

Storici, P., G. Capitani, et al. (1999). "Crystal structure of GABA-aminotransferase, a target for antiepileptic drug therapy." Biochemistry **38**(27): 8628-34.

Storici, P., G. Capitani, et al. (1999). "Crystal structure of human ornithine aminotransferase complexed with the highly specific and potent inhibitor 5-fluoromethylornithine." J Mol Biol **285**(1): 297-309.

Strisovsky, K., J. Jiraskova, et al. (2003). "Mouse brain serine racemase catalyzes specific elimination of L-serine to pyruvate." FEBS Lett **535**(1-3): 44-8.

Sugio, S., G. A. Petsko, et al. (1995). "Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity." Biochemistry **34**(30): 9661-9.

Sun, S. and M. D. Toney (1999). "Evidence for a two-base mechanism involving tyrosine-265 from arginine-219 mutants of alanine racemase." Biochemistry **38**(13): 4058-65.

Sundararaju Bakthavatsalam, A. A. A., Robert S. Phillips, Tatyana V. Demidkina, Maria V. Barbolina, Paul Gollnick, G. Guy Dodson, and Keith S. Wilson (1997). "The Crystal Structure of Citrobacter freundii Tyrosine Phenol-lyase Complexed with 3-(4'-Hydroxyphenyl)propionic Acid, Together with Site-Directed Mutagenesis and Kinetic Analysis, Demonstrates That Arginine 381 Is Required for Substrate Specificity." Biochemistry **36**(21): 6502-6510.

Tanase, S., H. Kojima, et al. (1979). "Pyridoxal 5'-phosphate binding site of pig heart alanine aminotransferase." Biochemistry **18**(14): 3002-7.

Thornton, J. M., A. E. Todd, et al. (2000). "From structure to function: approaches and limitations." Nat Struct Biol **7 Suppl**: 991-4.

Toney, M. D., E. Hohenester, et al. (1993). "Dialkylglycine decarboxylase structure: bifunctional active site and alkali metal sites." Science **261**(5122): 756-9.

Toney, M. D., S. Pascarella, et al. (1995). "Active site model for gamma-aminobutyrate aminotransferase explains substrate specificity and inhibitor reactivities." Protein Sci **4**(11): 2366-74.

Tsai, M. D., H. J. Weintraub, et al. (1978). "Conformation-reactivity relationship for pyridoxal Schiff's bases. Rates of racemization and alpha-hydrogen exchange of the pyridoxal Schiff's bases of amino acids." Biochemistry **17**(16): 3183-8.

Ura, H., K. Harata, et al. (2001). "Temperature dependence of the enzyme-substrate recognition mechanism." J Biochem (Tokyo) **129**(1): 173-8.

Vallee, B. L. and K. H. Falchuk (1993). "The biochemical basis of zinc physiology." Physiol Rev **73**(1): 79-118.

Vederas, J. C. (1980). "Stereochemistry of Pyridoxal Phosphate Catalyzed Enzyme Reactions." Accounts of Chemical Research **13**(12): 455-463.

Walsh, C., R. A. Pascal, Jr., et al. (1981). "Mechanistic studies on the pyridoxal phosphate enzyme 1-aminocyclopropane-1-carboxylate deaminase from Pseudomonas sp." Biochemistry **20**(26): 7509-19.

Webb, E. (1992). Enzyme Nomenclature 1992: Recommendations of the NCIUBMB on the Nomenclature and Classification of Enzymes.

Weber, I. T., L. N. Johnson, et al. (1978). "Crystallographic studies on the activity of glycogen phosphorylase b." Nature **274**(5670): 433-7.

Wilson, A. C., S. S. Carlson, et al. (1977). "Biochemical evolution." Annu Rev Biochem **46**: 573-639.

Wolfenden, R. (1974). "Enzyme catalysis: conflicting requirements of substrate access and transition state affinity." Mol Cell Biochem **3**(3): 207-11.

Wolosker, H., K. N. Sheth, et al. (1999). "Purification of serine racemase: biosynthesis of the neuromodulator D-serine." Proc Natl Acad Sci U S A **96**(2): 721-5.

Yano, T., S. Kuramitsu, et al. (1992). "Role of Asp222 in the catalytic mechanism of Escherichia coli aspartate aminotransferase: the amino acid residue which enhances the function of the enzyme-bound coenzyme pyridoxal 5'-phosphate." Biochemistry **31**(25): 5878-87.

Yano, T., S. Oue, et al. (1998). "Directed evolution of an aspartate aminotransferase with new substrate specificities." Proc Natl Acad Sci U S A **95**(10): 5511-5.

Chapter 2

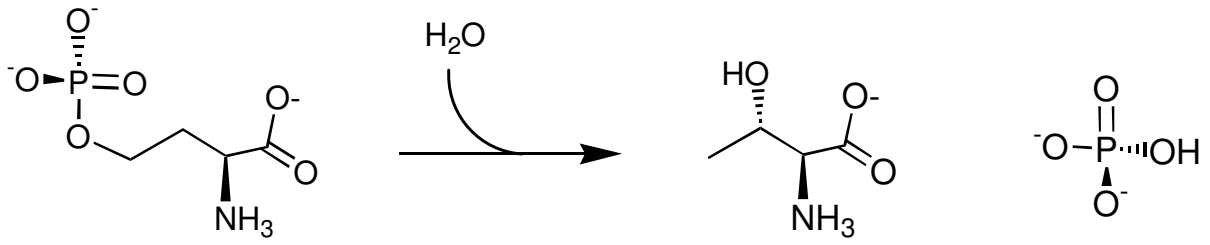
mTSh2: a threonine synthase homolog from mouse

Introduction

L-threonine (L-Thr) is known as an essential amino acid for many metazoans, including insects, birds and mammals (Kasting & McGinnis, 1958; Fitzgerald & Szmant, 1997; Reeds, 2000). 'Essential' means that these animals must obtain L-Thr from food sources, since their body is unable to synthesize it (Reeds, 2000). In contrast, plants and microorganisms can synthesize L-Thr endogenously, *via* a specific metabolic pathway (Kaplan & Flavin, 1965; Cami, 1993; Azevedo, 1997) the last step of which is catalyzed by threonine synthase (TS; EC 4.2.3.1).

TS is an enzyme that employs the organic cofactor pyridoxal-5'-phosphate (PLP) to carry out the conversion of *O*-phospho-L-homoserine (PHS) to threonine and phosphate (**Scheme 1**).

Scheme 1



O-phospho-L-homoserine L-Thr phosphate

Scheme 1. Threonine synthase reaction.

Structurally speaking, TS is a member of the fold-type II group of PLP-dependent enzymes (sometimes also termed β-family) (Grishin, 1995; Mehta & Christen, 2000; Christen & Mehta, 2001). This group encompasses enzymes that catalyze transformations of amino acids with covalency changes occurring not only at Cα but also at the β and/or γ carbons. Representative members of this group, in addition to TS, are tryptophan synthase, cystathionine β-synthase, *O*-acetylserine sulfhydrylase and L-serine ammonia-lyase (Mehta & Christen, 2000).

Recently our group has scanned a series of complete or near-complete genomes to obtain an inventory and a preliminary classification of genes that encode PLP-dependent enzymes (Percudani & Peracchi, 2003). This analysis identified, in

the genomes of several vertebrates (including man) two genes showing a high similarity to microbial TSs. This observation raised the intriguing possibility that threonine might be, after all, synthesized in humans and in other animals.

If L-Thr were produced, at least under some physiological or developmental conditions, this could have important biological implications. Alternatively, the existence of TS-like sequences in animal genomes may mean that the gene products have acquired new physiological roles upon dismissal of the threonine biosynthetic pathway during vertebrate evolution.

Herein, I describe a bioinformatic analysis of the two TS-like genes, as well as the cloning of a cDNA encoding a murine TS homolog and the characterization of the expressed and purified recombinant protein.

Materials and Methods

Materials

DNA primers were purchased from MWG Biotech (Ebersberg, Germany). The clone of homoserine kinase from *Methanococcus jannaschii*, inserted in an expression vector, was a kind gift from Hong Zhang (University of Texas Southwestern Medical Center). The recombinant enzyme was expressed and purified as described (Zhou, 2000; Krishna, 2001).

Rabbit muscle lactate dehydrogenase, L-homoserine, L-threonine and L-serine were from Fluka; *O*-phospho-L-threonine (PThr) and other potential TSh substrates (*O*-succinyl-homoserine, *O*-phospho-L-serine, etc.) were from Sigma-Aldrich. PHS was synthesized employing homoserine kinase, as described previously (Laber , 1994), and purified on an Amberlite IRA-410 anion-exchange column (Fluka). ^{32}P-labeled PHS was also prepared enzymatically, using homoserine kinase and γ-^{32}P-ATP.

The reaction products were run on a 20% polyacrylamide gel (30 min, 25 watts); the band corresponding to PHS was identified by autoradiography, cut and eluted in distilled water.

Bioinformatics

Initial identification of TS homologs (TSh) in mammalian genomes was achieved through a search of PLP-dependent enzymes in genomic sequences using profiles Hidden Markov Models (HMM; Percudani & Peracchi, 2003). Definition of the precise coding sequence boundaries of the matched genes was based on protein HMM and EST comparisons using Genewise and Estwise (http://www.ebi.ac.uk/Wise2); this procedure helped identification of the intron-exon structure of *TSh* genes and of alternative splice products.

The tissue-specific expression profile of mouse *TSh* sequences was inferred using the UniGene's EST ProfileViewer at the National Center for Biotechnology Information (NCBI). Homology searches were performed at the NCBI using Blast and genomic Blast. Multiple sequence alignments were produced with ClustalX (Thompson, 1997).

Phylogenetic analysis of TS and TSh proteins was performed with the maximum-likelihood method as implemented in the proml program of the PHYLIP package (http://evolution.genetics.washington.edu/phylip.html).

Recombinant expression and purification of the mouse TSh2 protein (mTSh2)

A clone of the complete cDNA for *mTSh2* (IMAGE Consortium ID 5250319, cDNA from mouse mammary tumor) was obtained from RZPD (Berlin, Germany).

The mTSh2 coding sequence was amplified from this clone, using the Pfu Turbo DNA polymerase (Perkin–Elmer) and adopting the following ''touchdown'' PCR conditions: 5 min of initial denaturation at 94°C, 10 cycles of 30 s denaturation at 94°C and 30s annealing at progressively lower temperatures (from 60 to 55°C, with a decrease of 1°C every other cycle), followed by 25 additional cycles at an annealing temperature of 55°C. The gene-specific primers were:

mTSh2_cpo+	taCGGTCCGatgtggtacactagcacccg
mTSh2_cpo-	taCGGACCGttccagcctgtctattctgag

Underlined residues correspond to the start codon and to the stop codon. The 5' region of both primers contained sequences (uppercase letters) recognized by the restriction enzyme *Cpo*I, so that the amplification product contained restriction sites near to both ends.

The amplicon was treated with *Cpo*I and directly cloned into the expression vector pET11-H6-*Cpo*. This plasmid is a derivative of pET11 (Novagen, Madison, WI) modified to present a single *Cpo*I restriction site in the cloning region, downstream to a sequence encoding a hexahistidine tag (Bolchi, 2005). The cloned insert was then sequence-verified and the plasmid was transformed into BL21-CodonPlus®-RIL *Escherichia coli* cells (Stratagene).

In these cells the recombinant protein was expressed efficiently, but under most induction conditions it was found exclusively in the insoluble fraction of the bacterial lysate. Attempts were made to improve the solubility of mTSh2 by reducing the growth temperature, varying the induction time and varying the concentration of added isopropyl β-D-thiogalactoside (IPTG).

A modest but sufficient yield of soluble protein (~3% of the total recombinant protein) was obtained by carrying out the induction with 0.1 mM IPTG and by allowing protein expression to proceed for 8 h at 24 °C.

After cell lysis, the soluble lysate fraction was loaded onto a metal-affinity resin (Talon; Clontech) equilibrated in Buffer A (5% glycerol/ 300 mM NaCl/ 50 mM phosphate-Na buffer pH 7.0) supplemented with 40 μM PLP, 5 mM β-mercaptoethanol and protease inhibitors (1 μM leupeptine, 1 μM pepstatine, 0.5 mM benzamidine, 0.5 mM phenyl methyl sulfonyl fluoride).

The recombinant mTSh2, bearing an N-terminal hexahistidine tag, was purified following the manufacturer's instructions.

The composition and purity of protein fractions were assessed by gel electrophoresis on SDS-polyacrylamide (11%) gels (Laemmli, 1970); mTSh2 fractions with a purity > 90% were pooled, dialyzed against buffer A supplemented with 10 μM PLP and 1 mM dithiothreitol, and stored at -80 °C.

The concentration of purified mTSh2 was estimated by the Bradford assay (Bradford, 1976) with bovine serum albumin as a standard, and confirmed based on the calculated extinction coefficient at 280 nm (60440 M^{-1} cm^{-1}; http://expasy.org/tools/protparam.html) (Gill & von Hippel, 1989).

Spectrophotometric measurements

Solutions containing recombinant mTSh2 (6 μM final) in 50 mM PIPES-NaOH buffer (pH 6.8), were supplemented with various potential substrates. Spectral changes were monitored using a Cary 400 spectrophotometer (Varian).

Cuvette holders were thermostated at 25 °C. Time-dependent spectral changes (time-courses) and titration data were analyzed by nonlinear least-squares fitting to the appropriate kinetic or thermodynamic equation using Sigma Plot (SPSS Inc.).

Enzymatic assays

The release of phosphate from PHS was established by incubating a trace of ^{32}P-labeled PHS (~1 nM) with mTSh2 (~1 μM) in 30 mM PIPES-Na buffer, pH 6.8 to 8.0, 25°C. Time-points were collected at appropriate intervals and quenched with 6 M urea. Substrates and products were separated on a 20% polyacrylamide gel and quantitated using a Personal Molecular Imager (Biorad). To make sure that the observed release of radioactive phosphate could be attributed to the recombinant protein, control experiments were carried out in which the enzyme was either omitted,

heat-denatured (by treatment at 100ºC for 5 min) or inactivated by reduction of the protein-bound PLP cofactor with sodium borohydride: in all these cases the reaction was essentially abolished.

The possible formation of threonine was tested by incubating PHS with mTSh2 for a few hours (30 mM PIPES-Na buffer, pH 6.8) and then running the reaction mixture on a silica TLC plate, side-by-side with amino acid standards (PHS, homoserine and L-Thr). Amino acids were visualized with ninhydrin (Friedman, 2004).

The rates of α-ketobutyrate production were measured spectrophotometrically by a coupled assay, in which mTSh2 (2 μM) and PHS (or PThr) were incubated in the presence of lactate dehydrogenase and NADH. The reaction mixture (200 μl final volume) contained 30 mM PIPES-Na (pH 6.8), 10 μM PLP, 0.25 mM NADH, 12.5 units/ml lactate dehydrogenase. The reactions were conducted at 25 ºC, or 37 ºC, and the disappearance of NADH was monitored at 340 nm.

Mass spectrometry identification of α-ketobutyrate

The recombinant mTSh2 protein (7.5 μM) was incubated with 8 mM PHS for a few hours, in a buffer solution (pH 7) containing 4 mM PIPES, 8 mM phosphate and 9.4 mM imidazole, plus 6 μM PLP. In parallel, a control reaction was conducted omitting the protein.

15 μl of the reaction mixture were diluted to 100 μl with 0.1 M formic acid. Non-charged solutes were then extracted with 100 μl of dichloromethane, and the organic phase was diluted to 1 ml with ethyl acetate. The sample was directly perfused into an ESI-MS spectrometer (Micromass ZMD 4000, equipped with a single quadrupole analyzer) at 20 μl min^{-1} rate (other working conditions: negative ion mode, capillary voltage 3.0 kV, cone voltage 30 V, source temperature 180 ºC, desolvation temperature 80 ºC).

In the mass spectrum of the enzymatic reaction mixture the base peak showed an m/z ratio of 101.2, identical within error to that expected for α-ketobutyrate (M-1 species). Such a peak was totally absent in the mass spectrum of the control reaction.

Results

Identification of the TSh genes

Our group is developing a curated database of PLP-dependent enzymatic activities, as well as bioinformatic tools for the identification and functional classification of PLP-dependent enzymes in genomic sets of predicted protein sequences. While performing an inventory of PLP-dependent enzymes in different organisms (Percudani & Peracchi, 2003) it was noticed that, in several vertebrate genomes, contained genes encoding proteins highly similar to TSs. These TS homologs (TShs) could be subdivided in two distinct types, which generally were co-present in the same organism.

A first group of genes encoded proteins (TSh1 proteins) that contain, in addition to the TS-like sequence, an N-terminal domain strongly similar to bacterial shikimate kinase.; a second group of gene products (TSh2 proteins) lack such an additional domain.

A list of TSh genes, some of which are annotated as threonine synthase-like in public databases, is presented in **Table 1**.

These genes were found only in vertebrates and in organisms belonging to basal Deuterostomia lineages, such as the urochordate *Ciona intestinalis* and the echinoderm *Strongylocentrotus purpuratus* (**Table 1**). Both *TSh1* and *TSh2* sequences are apparently missing in insects (*Drosophila, Anopheles or Apis*) and in *Caenorhabditis* genomes.

Examination of the ESTs databases indicated that mRNAs for both genes are present in tissues from their respective organisms, confirming that the genes are expressed. Based on the modest frequency with which transcripts are found in public EST libraries, expression of *TSh1* and *TSh2* appears to occur at low levels in mouse and human. In mouse, *TSh1* gene transcripts are observed predominantly in the brain and endocrine glands, whereas *TSh2* transcripts are most abundant in the kidney and liver.

The two *TSh* genes show very distinct intron-exon organizations. In most organisms, the *TSh1* gene is constituted by an uninterrupted coding sequence, over 2000-nucleotides long.

In contrast, *TSh2* genes contain multiple intervening sequences (e.g., both the human and the mouse gene include seven introns) and may thus undergo alternative splicing. In particular, the human *TSh2* gene (*hTSh2*) gives rise to several documented alternative transcripts (http://www.ebi.ac.uk/asd/), the longest of which encodes a protein significantly smaller than the other mammalian *TSh2* gene products (**Table 1**).

Table 1
Some metazoan genes encoding TS-like proteins

Group	Organism	GenBank/Ensembl Accession #	Predicted Protein Size (aa)[a]
TSh1	*Homo sapiens*	CAH70791	743
	Pan troglodytes	XP_521671	743
	Pongo pygmaeus	CAH92923	743
	Macaca fascicularis	Q9BH05	743
	Mus musculus	NP_808256	747
	Rattus norvegicus	AAH91213	745
	Bos taurus	XP_581830	748
	Canis familiaris	XP_535167	743
	Gallus gallus	CAG32314	680
	Danio rerio	XP_700684	702
	Tetraodon nigroviridis	CAG05472	680
	Fugu rubripes	SINFRUP00000155654	695
	Strongylocentrotus purpuratus	XP_791506	710
TSh2	*Homo sapiens*	AAH64423	405
	Pongo pygmaeus	CAH89511	484
	Mus musculus	AAH51244	483
	Rattus norvegicus	NP_001009658	485
	Bos taurus	XP_615043	488
	Canis familiaris	XP_532965	484
	Gallus gallus	XP_420866	521
	Xenopus laevis	AAH84269	472
	Danio rerio	AAI10122	476
	Strongylocentrotus purpuratus	XP_793160	461
	Ciona intestinalis	AK115020	474

Table 1. The group, organism source, databank accession number, protein size, are shown for all the metazoan genes encoding threonine synthase-like proteins. [a]When multiple splice variants of the gene product are included in the databases, the size refers to the longest predicted polypeptide.

Comparison and phylogeny of TSh gene products

BLAST searches of the GenBank non-redundant database confirmed that *bona fide* TSs are the enzymes most similar to TShs. For example, when the mouse TSh2 (mTSh2) sequence was used as a query, the first match corresponding to a functionally validated protein was *Saccharomyces cerevisiae* TS (Ramos & Calderon, 1994) (38% identity to mTSh2 over 458 positions; E-value=10^{-80}), followed by several TSs from eubacteria, such as the enzyme from *Pseudomonas aeruginosa* (Clepet, 1992) (36% identity; E-value=2×10^{-77}).

For both TSh1 and TSh2 proteins, all these best matches belong to the so-called subfamily II of TSs, which includes monomeric proteins from fungal and eubacterial sources (Laber, 1999; Omi, 2003). A lesser similarity is found to TSs of subfamily I (enzymes that function as homodimers and occur in plants and archaea, as well as in some Gram-positive eubacteria).

The genes for TSh1 and TSh2 proteins do not seem, however, derived from a recent duplication of a single ancestor gene. In fact, it is noteworthy that TSh1 and TSh2 sequences are more similar to microbial TSs than they are to each other (only 26% identity is observed between the two mouse TSh sequences). Consistent with this hypothesis, phylogenetic analyses of TS and TSh sequences (**Figure 1**) indicate that the TSs of subfamily II fall into at least two distinct clades. Each clade contains both microbial and metazoan sequences, with one of the two TSh families at the higher end of the phylogenetic tree.

However, with respect to sequences belonging to the same clade, the metazoan TShs show some distinctive features. TSh1 sequences are characterized, as mentioned above, by an N-terminal kinase domain that is unique to this group of proteins, while TSh2 sequences contain mutations at some sites that are conserved in all functionally validated TSs.

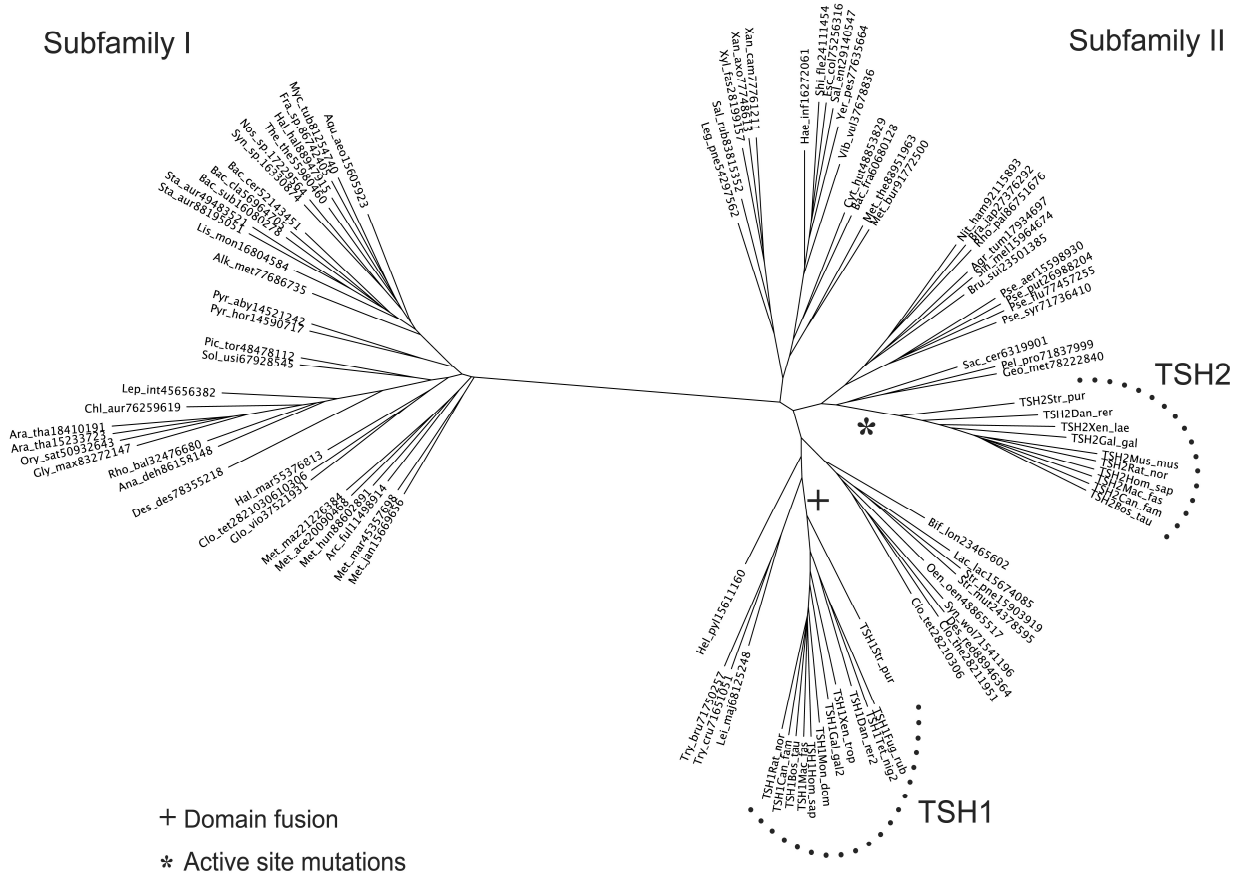


Figure 1. Phylogenetic relationships of TS and TSh proteins. The unrooted maximum likelihood phylogenetic tree was constructed with the proml program of the PHYLIP package and displayed with the TreeIllustrator program with branch length correction for visibility enhancement. In the case of TSh1 sequences, only the PLP-dependent domain was used for the phylogenetic reconstruction. Sequences are indicated with the abbreviation of the binomial name of species followed by the GenBank gi number. Terminal nodes corresponding to TSh1 and TSh2 sequences are indicated by dotted lines. A plus sign and an asterisk sign mark ancestral modifications leading to the TSh1 and TSh2 lineages, respectively.

The latter point is illustrated in **Figure 2**, which shows a multiple sequence alignment of the two mouse TShs with two validated TSs – the enzymes from *Thermus thermophilus* (subfamily I) and from yeast (subfamily II). At least 24 residues (boxed in the figure) are strictly invariant in all characterized TSs (Ramos & Calderon, 1994; Clept, 1992; Laber, 1999; Omi, 2003; Laber, 1994; Malumbres, 1994; Malumbres, 1995; Curien, 1996). All of these residues are conserved in the TSh1 protein lineage, but four of them are mutated in most mTSh2 proteins (**Figure 2**). The mutations, referred to the numbering of the *T. thermophilus* TS, are: Ala83Gly, Phe134Ser, Asn188Ala (in non-mammalian sequences: Asn188Gly) and Thr317Pro (or Thr317Ser).

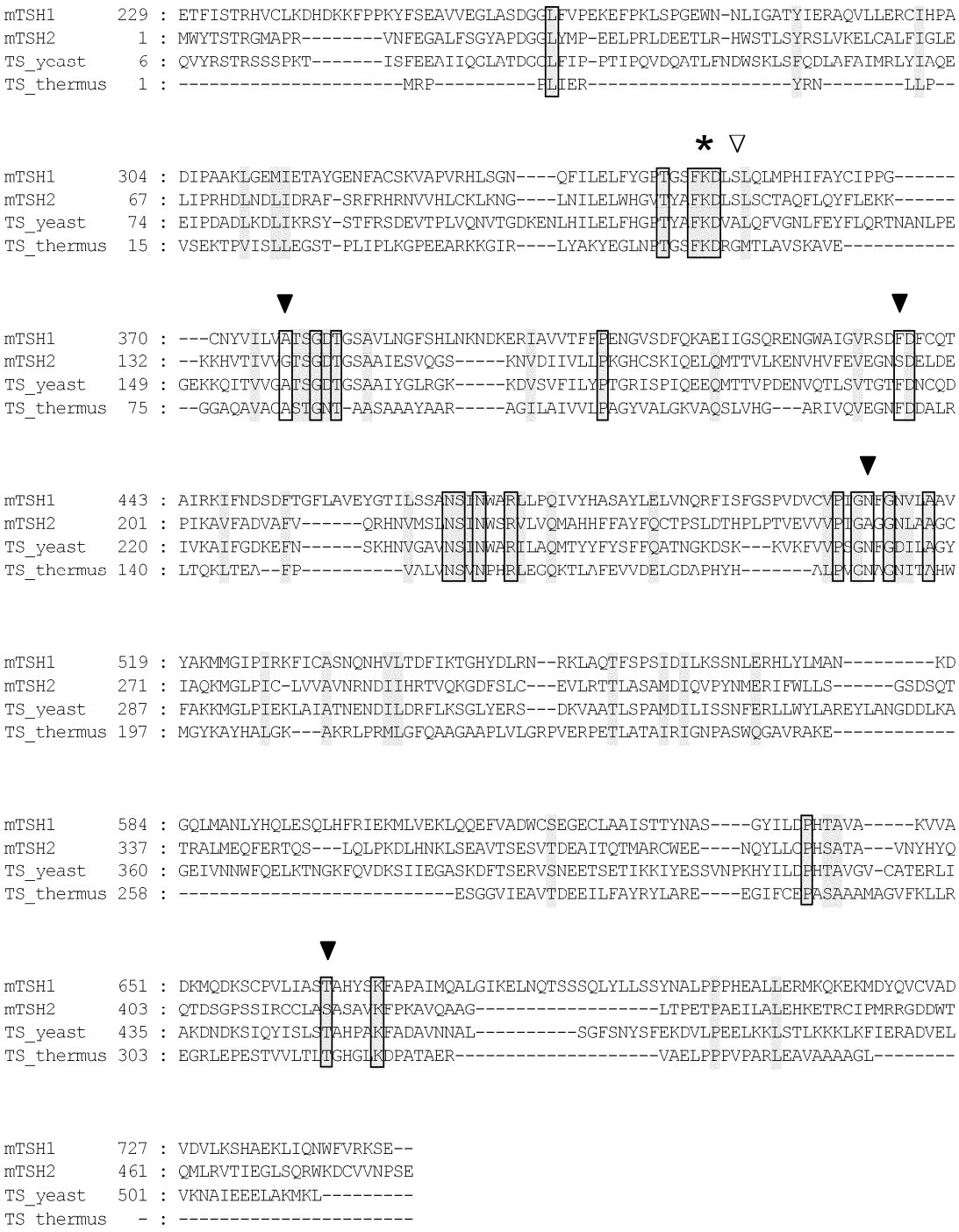


Figure 2. Multiple sequence alignment of the mouse TSh proteins with TSs from *Saccharomyces cerevisiae* (GenBank P16120) and from *Thermus thermophilus HB8* (GenBank BAD70314). For mTSh1, only the PLP-dependent domain is shown. Positions that are absolutely conserved in all functionally validated threonine synthases (but not necessarily in TSh proteins) are boxed. The black arrows signal those conserved residues that are mutated in mTSh2. The asterisk marks the catalytic lysine residue that binds the PLP cofactor through a Schiff base linkage. The white arrow indicates a position where all mammalian TSh1 and TSh2 proteins contain a Ser residue in place of Ala or Gly (as present in all TS). The mutations prevent the recognition of these proteins by the Prosite pattern for threonine synthases/dehydratases PLP attachment site (PS00165; http://www.expasy.org/prosite/).

Cloning, expression and purification of recombinant mTSh2

We obtained the cDNA clones corresponding to the full-length human *TSh1* (*hTSh1*) and mouse *TSh2* (*mTSh2*) genes from the IMAGE consortium (http://image.llnl.gov). Since *Escherichia coli* is currently the system of choice for the heterologous expression of bacterial and eukaryotic TSs (Laber, 1999; Omi, 2003; Garrido-Franco, 2002), this organism was used for the expression of recombinant TShs.

Several attempts to subclone the *hTSh1* gene in an *E. coli* expression vector were unrewarding: in the very few positive colonies, the cloned insert contained frameshifts or large deletions, suggesting that perhaps the intact hTSh1 protein is toxic to the host cells, even when expressed at very low levels in the absence of induction.

In contrast, overexpression of the mTSh2 protein was accomplished promptly, even though soluble protein could only be obtained in low yields. Approximately 0.4 mg of pure, recombinant mTSh2 protein was isolated per liter of bacterial culture (see Methods).

The recombinant protein was 506 residues long (including the N-terminal hexahistidine tag) with a calculated molecular mass of 56.5 kDa. Purified mTSh2 was homogenous by SDS-PAGE (data not shown), and exhibited the characteristic PLP absorption maximum at 410 nm (**Figure 3A**).

Interaction of mTSh2 with potential substrates

In PLP-dependent enzymes, the presence of a light-absorbing cofactor in the active site provides a convenient means for monitoring the binding of substrates, substrate analogs and inhibitors, as well as the formation of catalytic intermediates (e.g., Laber, 1994; Mozzarelli, 1989; Sterk & Gehring, 1991; Karsten & Cook, 2002).

The interaction of mTSh2 with PHS (the physiological substrate of TSs) have been tested spectroscopically, with L-Thr and with the following compounds that show similarity to either the substrate, the product or to intermediates of the TS reaction: L-homoserine, *O*-succinyl-homoserine, 2-amino-5-phosphonopentanoate (AP5), PThr, L-*allo*-threonine, *O*-phospho-L-serine (Pser), *O*-phosphoethanolamine, L-vinylglycine, cystathionine, L-homocysteate and *O*-phospho-L-tyrosine. Addition of these compounds (millimolar concentrations) to the enzyme led in most cases to negligible changes in the PLP spectrum. L-Thr and L-allo-threonine induced a slight increase and a red shift of the main PLP band to 414 nm (data not shown), hinting that the enzyme may be binding these amino acids, albeit with low affinity.

However, much more striking spectral changes were observed with PHS, AP5 and PThr (**Figure 3A**).

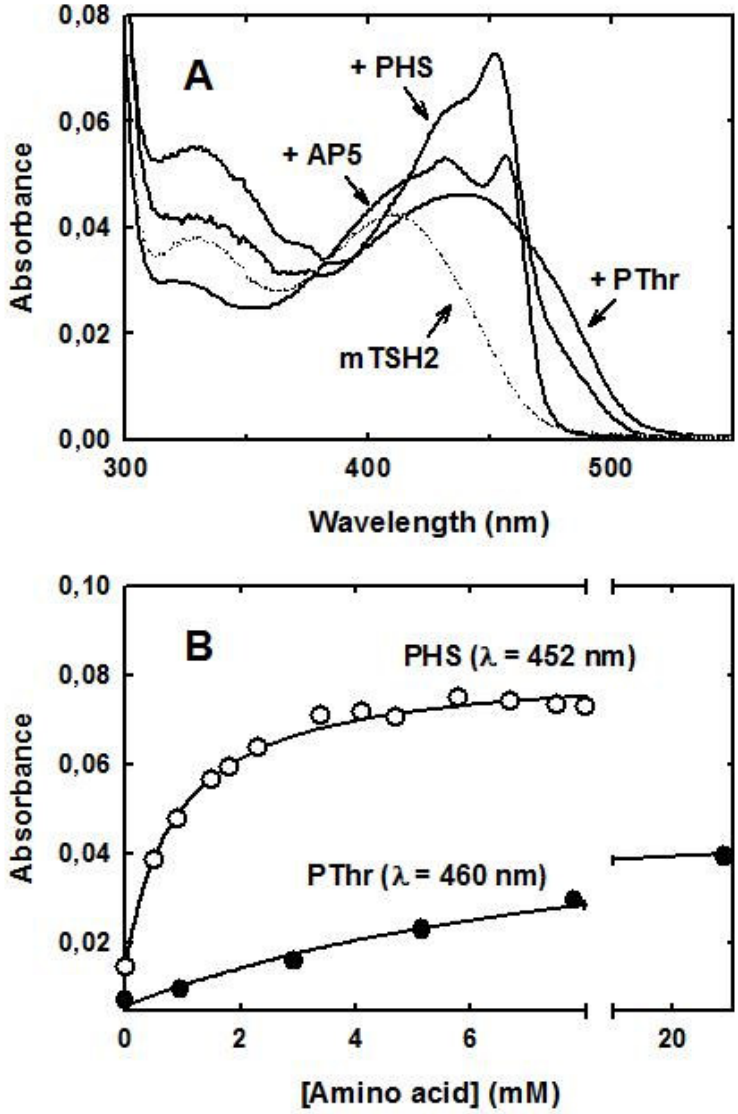


Figure 3. (A) Absorption spectra of recombinant mTSh2 in the presence and absence of amino acid ligands. The protein (6 μM) was incubated in 50 mM PIPES-NaOH, pH 6.8, 25°C. Absorption spectra were recorded in the absence of ligands (dotted line) or in the presence of 5 mM PHS or 20 mM PThr or 20 mM AP5. (B) Spectroscopic titration of the mTSh2 protein with PHS and PThr. Solid lines are best fits to binding isotherms, yielding apparent K_d values of 0.9 mM (PHS) and 9.5 mM (PThr).

PHS induced the appearance of a sharp band at 452 nm, which decayed over time (hours). Such a band strongly resembled the peak that is observed upon the reaction of the *E. coli* TS with PHS (Laber, 1994) and that is attributed to a stabilized carbanionic intermediate (quinonoid) formed upon removal of the amino acid α-proton (Laber, 1994).

A titration with PHS yielded an apparent K_d of 0.9 mM (**Figure 3B**), comparable to the K_m of *E. coli* TS towards the same substrate (K_m=0.5 mM; Laber, 1994).

AP5, a well-known TS substrate analog (Laber, 1994), was also bound by mTSh2. Reaction of mTSh2 with racemic AP5 yielded a stable spectroscopic species, characterized by maxima at 432 and at 456 nm (**Figure 3A**). The latter peak, again attributable to a quinonoid adduct, is reminiscent of the peak that bacterial TS forms with the L-enantiomer of AP5 (Laber, 1994).

In contrast to the cases above, reaction of mTSh2 with PThr produced a broad band centered at ~450 nm, with a shoulder at 480 nm, and a substantial absorbance increase at 330 nm (**Figure 3A**). This suggests that PThr binding leads to a mixture of PLP-bound species, but not to the accumulation of a quinonoid intermediate (apparent $K_d \approx 10$ mM; **Figure 3B**).

mTSh2 catalyzes slow phospho-lyase reactions with PHS and PThr

We studied in more detail the reaction of mTSh2 with PHS. Reactions carried out with ^{32}P-5'-labeled PHS showed that mTSh2 was capable of releasing the phosphate group from the amino acid, albeit at a modest rate (**Figure 4**). Phosphate release is of course accomplished also by *bona fide* TSs.

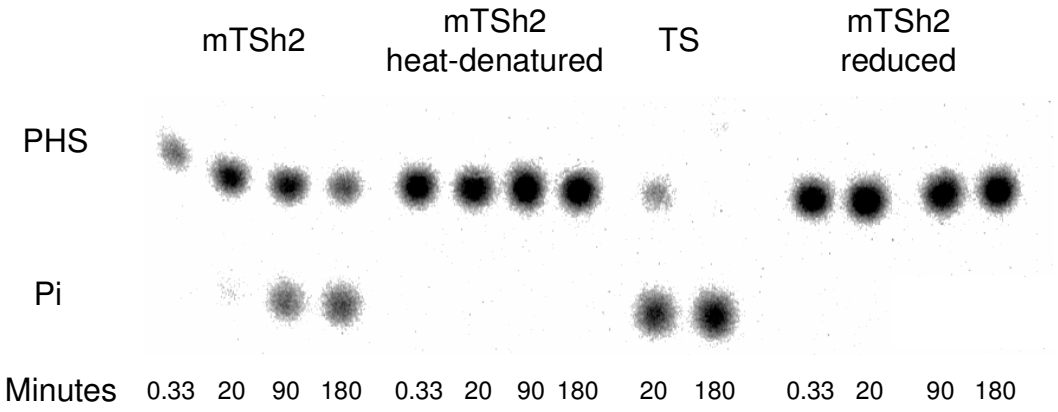


Figure 4. Analysis of phospho-lyase activity of mTSh2. Individual reactions were realized as described in the Materials and Methods. mTSh2 represent the reaction with the enzyme, mTSh2 heat-denatured and mTSh2 reduced are the negative; TS is the positive control (TS is a cellular extract of a strain of *Escherichia coli* overexpressing Threonine synthase). Reactions were incubated at 25°C in 30 mM PIPES-Na buffer, pH 6.8. Time-points were collected at 0.33, 20, 90 and 180 minutes, for reactions with mTSh2 and the negative controls, while for the positive control time-points were collected at 20 and 180 minutes. PHS and Pi were separated by denaturing PAGE, and results were quantitated using a Phosphor-Imager.

However, when the recombinant mTSh2 was incubated with PHS for a few hours and the reaction products were analyzed by TLC followed by ninhydrin coloration, no evidence was found for the generation of L-Thr (or of homoserine; data not shown). While

the physiological role of TS is to convert PHS to L-Thr and phosphate (formally a γ,β-elimination/replacement reaction), the bacterial enzyme also slowly catalyzes a parasitic phospho-lyase (α,γ-elimination) reaction, converting PHS to α-ketobutyrate, phosphate and ammonia (Schildkraut & Greer, 1973). Two evidences indicated that mTSh2 was also capable of converting PHS to α-ketobutyrate.

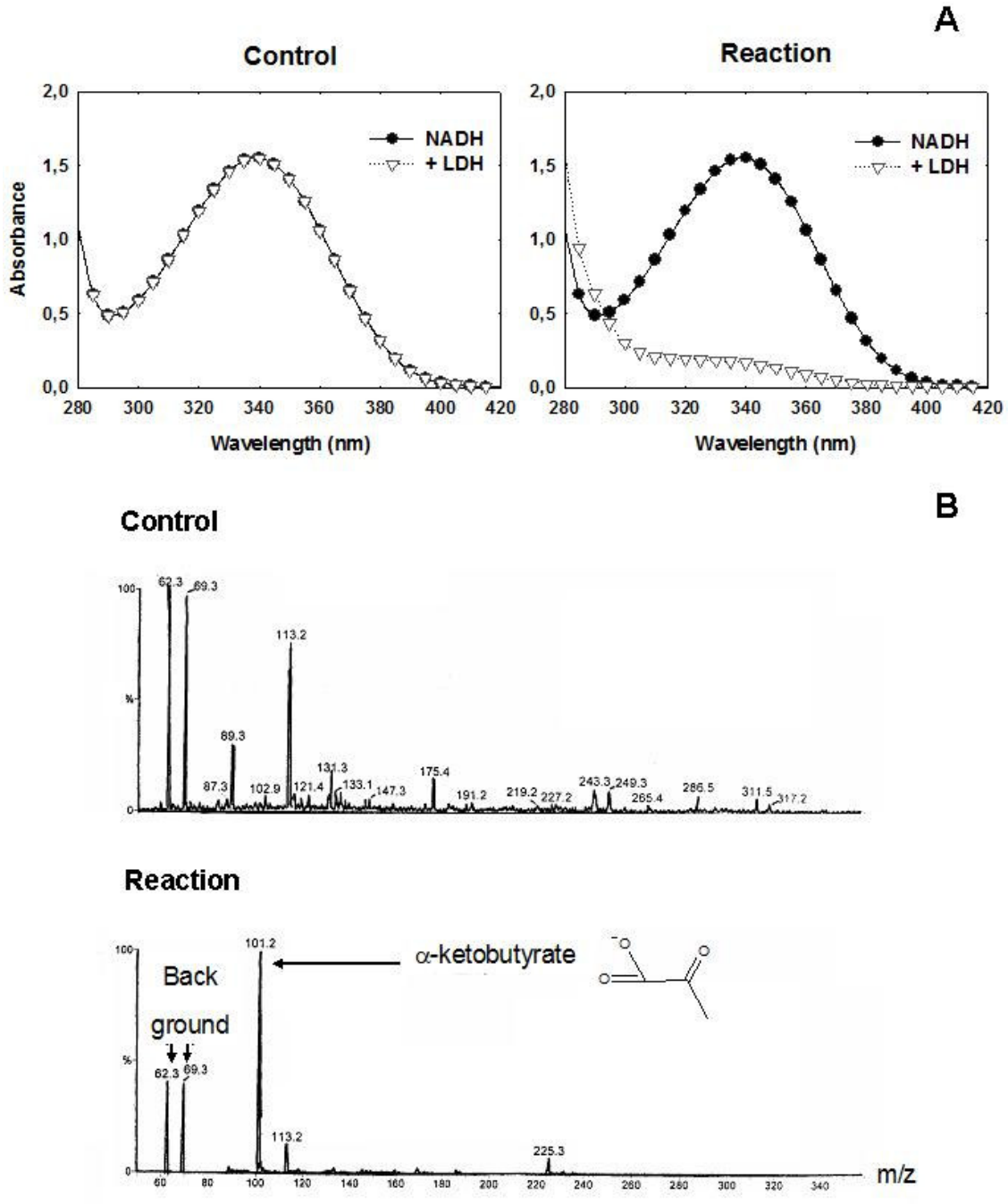


Figure 5. Analysis of reaction products. Both Control and Reaction were incubated 8 hours at 25°C in 30 mM PIPES-Na buffer, pH 6.8. In the control the enzyme was omitted. (A) The presence of α-ketoacid in the reaction mixture was evaluated by measuring the oxidation of NADH at 340 nm (see Materials and Methods) after the addition of Lactate dehydrogenase. (B) ESI mass spectrum of the reaction mixture. The base peak shows an m/z ratio of 101.2 has reference to the expected product α-ketobutyrate.

First, when the products of the reaction between mTSh2 and PHS were added to a solution containing lactate dehydrogenase and NADH, they induced a complete oxidation of NADH to NAD^+, suggesting that these products included a short-chain α-ketoacid (**Figure 5A**).

Second, an analysis of the reaction products by electrospray ionization mass spectrometry (negative mode) revealed an intense peak corresponding to the mass expected for α-ketobutyrate (**Figure 5B;** see Materials and Methods).

TSs are capable of forming α-ketobutyrate also from L-Thr (Laber , 1994; Skarstedt & Greer, 1973), but we exclude that this mechanism might be operating in the case of mTSh2. In fact, as noted, L-Thr was not formed in detectable amounts upon incubating mTSh2 with PHS. Moreover, the kinetics of formation of α-ketobutyrate, measured through a coupled assay with lactate dehydrogenase, were linear with time, with no lag phases. Finally, when mTSh2 was incubated with an excess L-Thr (10 mM), no appreciable formation of α-ketobutyrate was detected in several hours. Thus, α-ketobutyrate must be formed directly from PHS. The putative mechanism for such an α,γ-elimination reaction is given in **Figure 6**.

The rate constant of α-ketobutyrate formation, measured in the presence of saturating PHS (6 mM), was 0.03 s^{-1} (pH 6.8, 25°C) and somewhat slower at higher pH values (up to pH 8.0; data not shown). When the kinetics were measured under near-physiological conditions (pH 7.4, 37°C; **Figure 7A**) the calculated maximum activity was ~80 nmoles min^{-1} mg^{-1}, with a K_M of 0.42 mM.

Interestingly, the recombinant protein was also capable of producing α-ketobutyrate from PThr, at about the same rate (0.04 s^{-1} in the presence of 20 mM PThr), while it was almost inactive towards PSer. This datum indicates that mTSh2 can carry out both α,γ-eliminations and α,β-eliminations on phosphorylated substrates.

The PThr phospho-lyase reaction was somewhat faster than the reaction with PHS, but required higher substrate concentrations (**Figure 7B**). In fact, the calculated maximum activity of recombinant mTSh2 was ~500 nmoles min^{-1} mg^{-1}, with a K_M of 12 mM, with PThr as substrate. Not unexpectedly, the reaction was inhibited competitively by phosphate (**Figure 7B)**.

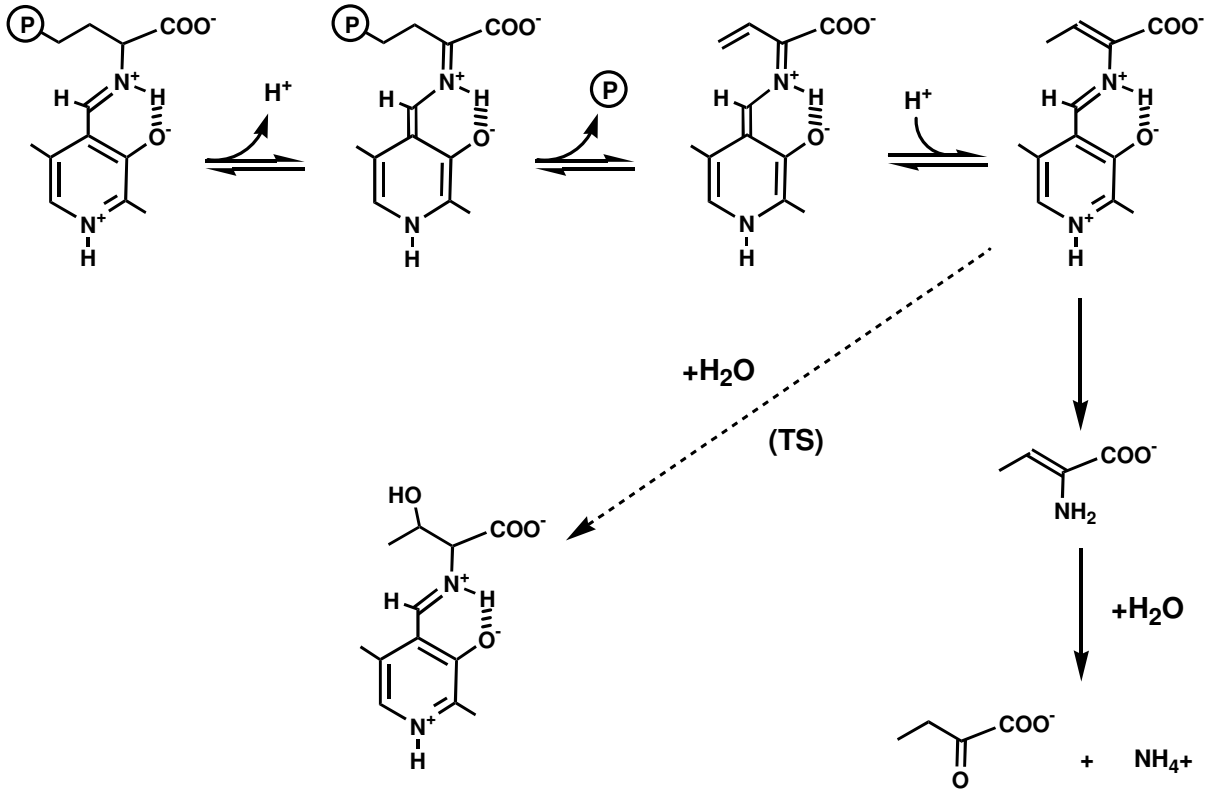


Figure 6. Proposed mechanism for the reaction catalyzed by mTSh2 with PHS. The substrate (PHS) binds initially to the cofactor forming an aldimine (external Schiff base; the phosphate group of PLP is omitted for clarity). Removal of the α-proton of the amino acid yields a stabilized carbanionic intermediate (quinonoid, characterized by an absorption maximum in the 450-500 nm region). Subsequently, elimination of the phosphate group and protonation of the γ-carbon leads to formation of a new aldimine, the Schiff base of aminocrotonate. In the reaction catalyzed by TSs, the double bond of this intermediate undergoes a stereospecific hydration, to generate the Schiff base of L-Thr (dashed arrow). In mTSh2, instead, aminocrotonate may be released into the solvent, where it readily hydrolyzes to α-ketobutyrate and ammonia.

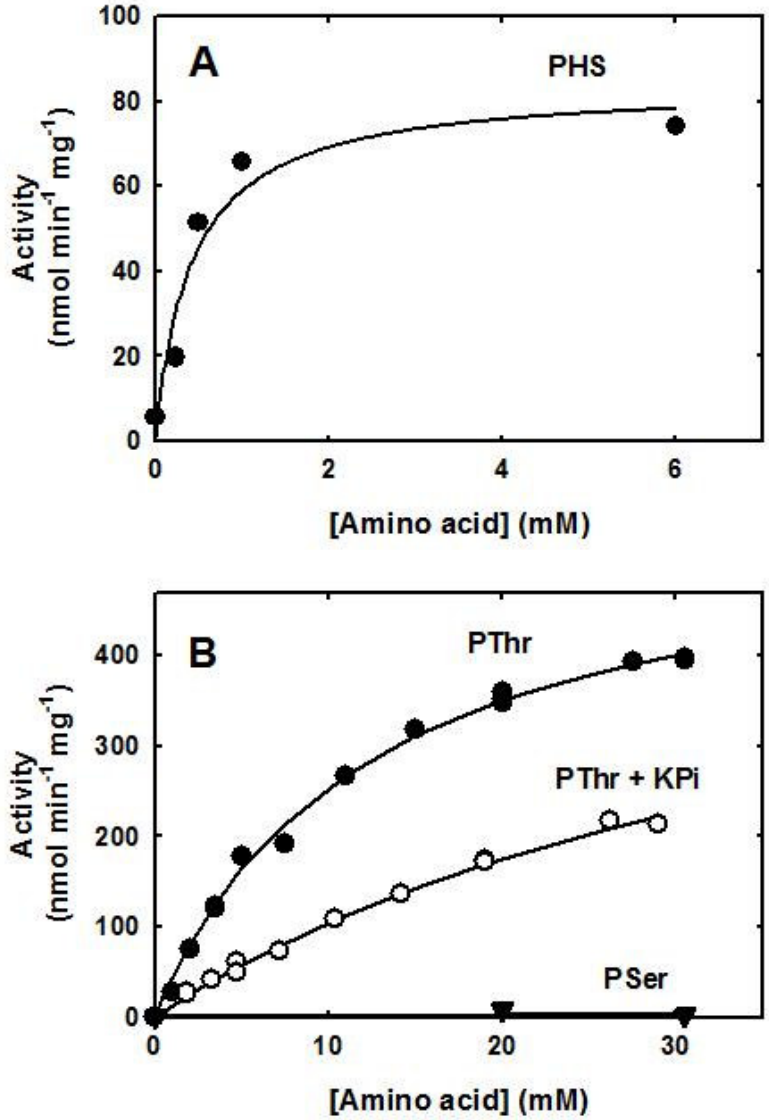


Figure 7. Kinetics of α-ketobutyrate formation. Initial rates of product formation were measured at pH 7.4, 37°C, in the absence or in the presence of 10 mM potassium phosphate. Least-squares fits to the Michaelis-Menten equation yielded. (A) PHS phospho-lyase reaction. V_{max}=84 nmoles min^{-1} mg^{-1} and K_M=0.4 mM (no phosphate). (B) PThr phospho-lyase reaction. V_{max}=560 nmoles min^{-1} mg^{-1} and K_M=12 mM (no phosphate, black circles) and V_{max}=540 nmoles min^{-1} mg^{-1} and K_M=41 mM (10 mM phosphate, white circles). The phospho-lyase reaction with PSer was ≥100-fold slower (no phosphate, black triangles down).

Discussion

Why mTSh2 is not a TS

mTSh2 shows a 38% identity to the yeast TS, but lacks TS activity, confirming that sequence similarity is a weak criterion for ascribing a function to a newly identified enzyme (Rost, 2002). A finer analysis of conservation at key active site positions helps understand why mTSh2 does not act as a TS.

There are a few crystal structures of TSs, showing the arrangement of amino acids in the active site (Omi, 2003; Garrido-Franco, 2002; Mass-Droux, 2006). In particular, a structure of the *T. thermophilus* TS (tThrS) complexed with the PHS analogue AP5 shows the interactions of the PLP-ligand conjugate with the active site residues (Omi, 2003), providing insights into TS function. In this structure, Phe134 and Asn188 interact closely with the side chain of the substrate analogue: the Phe ring forms a van der Waals contact with the γ-carbon of AP5, whereas the Asn amido group donates a hydrogen bond to the phosphonate of AP5 (Omi, 2003; **Figure 8**).

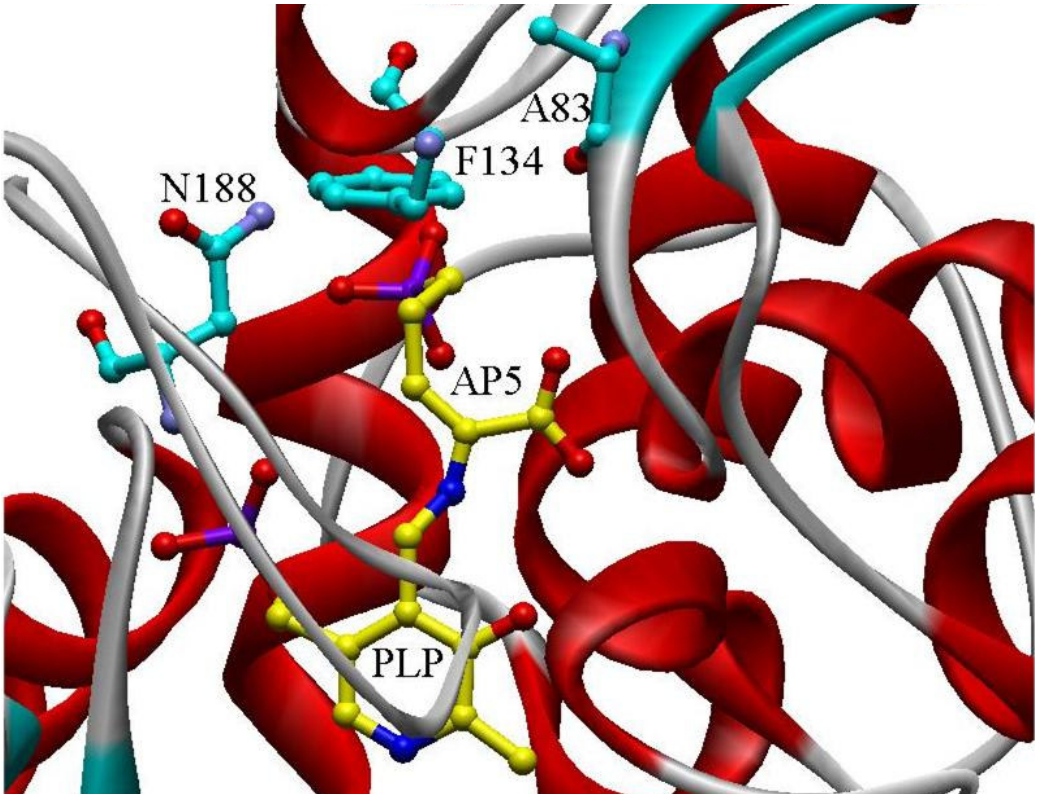


Figure 8. Stereoview of the active site in tThrS**.** Close-up view of the active site of the tThrS-AP5 complex in the closed form. The front and back of the figure are the solvent side (entrance of the active site) and the protein side (bottom of the active site), respectively. The α-helices, β-strands and loops are drawn in *red*, *light blue, gray.* Ala^{83}, Phe^{134}, Asn^{188} are coloured in *light blue*. The cofactor and AP5 are drawn in *yellow* with *red phosphates*.

In mTSh2, Phe134 and Asn188 are replaced by smaller amino acids (Ser and Ala, respectively; **Figure 2**), which seems consistent with a significant divergence in reaction specificity or mechanism. For example, since Phe134, Asn188 and Ala83 are located near the entrance of the TS active site, the mTSH2 active site could be more open to the solvent.

An enzyme with phospho-lyase activity

The occurrence of several active-site mutations might even suggest that mTSh2 is catalytically inactive. Recent genomic observations indicate that, in metazoa, enzyme homologs often assume non-catalytic functions – a role switch signaled by the accumulation of inactivating mutations (Pils & Schultz, 2004). In contrast to this hypothesis, however, we have shown that mTSh2 can still bind PLP, react with potential substrates (including PHS) forming recognizable reaction intermediates, and carry out specific elimination reactions.

What is the physiological function of mTSh2? The preferential transcription of the gene in the liver and kidney is consistent with a catabolic role for the enzyme. The selectivity of recombinant mTSh2 toward a battery of ligands shows a preference for large phosphorylated amino acids. Both observations are substantiated by the finding that mTSh2 catalyzes elimination reactions with PHS and PThr. Nevertheless, it is doubtful that the reaction with PHS may have any metabolic relevance, as this compound is not known to be formed in animal metabolism (Connick, 1986). Free PThr, on the other hand, occurs in animal tissues, possibly deriving from the hydrolysis of phosphorylated proteins (Kataoka, 1991), and degradation of this amino acid may be one biological task of mTSh2.

The activity of mTSh2 in the PThr phospho-lyase reaction is relatively modest, but comparable to those of other PLP-dependent enzymes. For example, TS from *Arabidopsis thaliana*, in the absence of the allosteric activator S-adenosylmethionine, shows a similar turnover number (Curien, 1998). In mouse cells, mTSh2 might be activated allosterically or through prost-translational modifications; alternatively, the main physiological substrate for mTSh2 might be some phosphorylated amino compound other than PThr.

At least two PLP-dependent mammalian phospho-lyases have been described in the past: *O*-phosphoethanolamine phospho-lyase (Fleshood & Pitot, 1970) and *O*-phosphohydroxylysine phospho-lyase (Tsai & Henderson, 1974). Both were originally purified from rat liver, and both have remained uncharacterized at the molecular level.

Clearly, mTSh2 cannot be identified with *O*-phosphoethanolamine phospho-lyase, since it does not bind *O*-phosphoethanolamine and since its predicted mass (≈54 kDa for the nonrecombinant protein) does not fit the mass reported for the rat enzyme (168 kDa; Fleshood & Pitot, 1970), even considering the possibility that *in vivo* mTSh2 may form dimers, as is the case with some TSs.

The mass discrepancy also holds when comparing mTSh2 to *O*-phosphohydroxylysine phospho-lyase (140 kDa; Tsai & Henderson, 1974); however we did not directly test the activity of mTSh2 against *O*-phosphohydroxylysine, a compound not available from commercial sources.

Evolution of TSh2: a case of 'enzyme recycling'

During evolution, the development of new enzyme functions is believed to occur mainly by recruiting proteins that already exist and catalyze similar reactions (O'Brien & Herschlag, 1999). In the simplest scenario, the triggering event is a duplication of the gene encoding the enzyme, which relieves the selective constraints on one gene copy, allowing for potential functional diversification.

The TSh2 case hints at the possibility of a different evolutionary mechanism. Namely, it suggests that dismissal of the L-Thr biosynthetic pathway (somewhere after the separation between the fungal and animal lineages), rather than gene duplication, may have freed the TS gene from the selective constraints of ancestral function, allowing evolution of the gene product into a catabolic phospho-lyase.

Analogous instances of 'enzyme recycling' are rare, but not unprecedented. For example, while mammals have lost the ability to degrade allantoin, they retain some of the enzymes in the allantoin degradation pathway, for reasons that are not fully clear (Fujiwara & Noguchi, 1995; Vigetti, 2002).

Of course, we cannot rigorously exclude that TSh2 might have originally emerged as a duplicated copy of TS, before disappearance of the latter gene in the metazoan lineage. However, homology searches in complete genomes did not reveal any extant organism containing both a TS and a TSh2 gene (not shown).

It should be noted that the biochemical requisites underlying the 'enzyme recycling' mechanism sketched above are similar to those required by the 'gene duplication' scenario. In both instances, evolution of a novel catalytic function is greatly facilitated if the original enzyme already possesses (perhaps at low levels) a secondary activity, which can

be improved and selected (O'Brien & Herschlag, 1999). This is the case for TS, which is intrinsically able to act as a degradative phospho-lyase (Schildkraut & Greer, 1973).

Speculative considerations on TSh1

Compared to TSh2 proteins, the enzymes encoded by TSh1 genes show a slightly lower overall similarity to fungal TSs, but retain all the strictly conserved residues (**Figure 2**). Despite this, we deem that these enzymes, too, do not act physiologically as TSs. This view is in keeping with the long-standing recognition of L-Thr as an essential amino acid in animals (Fitzgerald & Szmant, 1997; Payne & Loomis, 2006) and with the absence of other L-Thr biosynthetic genes in vertebrate genomes (Payne & Loomis, 2006).

At present, we can only speculate about the possible function of TSh1 genes. In particular, since transcription of such genes occurs mainly in the brain and in endocrine glands, it is tempting to hypothesize the involvement of TSh1 in the metabolism of some chemical mediator. In fact, higher eukaryotes often employ PLP-dependent enzymes for the synthesis and degradation of hormones and chemical messengers, and these enzymes can be strict homologs of enzymes involved in basic metabolic pathways (Percudani & Peracchi, 2003; Iyer, 2004).

Conclusions

In the past, TS was considered a promising target for the development of selective antimicrobial drugs or herbicides (Harde , 1994), as it was postulated that this enzyme would not occur in animals, but only in prokaryotes, fungi and plants. However, the occurrence in vertebrates of TShs (at least some of which can bind the true TS substrate) casts serious doubts on this axiom, implying that the design of selective TS inhibitors may be extremely challenging.

Acknowledgments

This work was funded by the Italian MIUR (COFIN 2005) and by an EMBO YIP grant (to A.P.). I would like to thank H. Zhang for the HK clone and B. Laber for suggestions about PHS preparation. I would also like to thank A. Mozzarelli and S. Ottonello for support; and A. Bolchi, D. Lazzaretti, L. Ferrarini and F. Sansone for technical assistance.

References

Azevedo, R. A., P. Arruda, et al. (1997). "The biosynthesis and metabolism of the aspartate-derived amino acids in higher plants." Phytochemistry 46(3): 395-419.

Bolchi, A., S. Ottonello, et al. (2005). "A general one-step method for the cloning of PCR products." Biotechnol Appl Biochem 42(Pt 3): 205-9.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal. Biochem. 72: 248-254.

Cami, B., C. Clepet, et al. (1993). "Evolutionary comparisons of three enzymes of the threonine biosynthetic pathway among several microbial species." Biochimie 75(6): 487-495.

Christen, P. and P. K. Mehta (2001). "From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Chem. Rec. 1(6): 436-447.

Clepet, C., F. Borne, et al. (1992). "Isolation, organization and expression of the Pseudomonas aeruginosa threonine genes." Mol. Microbiol. 6(21): 3109-3119.

Connick, J. H., G. C. Heywood, et al. (1986). "O-phosphohomoserine, a naturally occurring analogue of phosphonate amino acid antagonists, is an N-methyl-D-aspartate (NMDA) antagonist in rat hippocampus." Neurosci. Lett. 68(2): 249-51.

Curien, G., R. Dumas, et al. (1996). "Characterization of an Arabidopsis thaliana cDNA encoding an S-adenosylmethionine-sensitive threonine synthase. Threonine synthase from higher plants." FEBS Lett. 390(1): 85-90.

Curien, G., D. Job, et al. (1998). "Allosteric activation of Arabidopsis threonine synthase by S-adenosylmethionine." Biochemistry 37(38): 13212-21.

Fitzgerald, L. M. and A. M. Szmant (1997). "Biosynthesis of 'essential' amino acids by scleractinian corals." Biochem. J. 322: 213-21.

Fleshood, H. L. and H. C. Pitot (1970). "The metabolism of O-phosphorylethanolamine in animal tissues. I. O-phosphorylethanolamine phospho-lyase: partial purification and characterization." J. Biol. Chem. 245(17): 4414-20.

Friedman, M. (2004). "Applications of the ninhydrin reaction for analysis of amino acids, peptides, and proteins to agricultural and biomedical sciences." J. Agric. Food Chem. 52(3): 385-406.

Fujiwara, S. and T. Noguchi (1995). "Degradation of purines: only ureidoglycollate lyase out of four allantoin-degrading enzymes is present in mammals." Biochem. J. 312: 315-8.

Garrido-Franco, M., S. Ehlert, et al. (2002). "Structure and function of threonine synthase from yeast." J. Biol. Chem. 277(14): 12396-405.

Gill, S. C. and P. H. von Hippel (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Anal. Biochem. 182(2): 319-26.

Grishin, N. V., M. A. Phillips, et al. (1995). "Modeling of the spatial structure of eukaryotic ornithine decarboxylases." Protein Sci. 4(7): 1291-1304.

Harde, C., K. H. Neff, et al. (1994). "Syntheses of homoserine phosphate analogs as potential inhibitors of bacterial threonine synthase." Bioorg Med. Chem. Lett. 4(2): 273-278.

Iyer, L. M., L. Aravind, et al. (2004). "Evolution of cell-cell signaling in animals: did late horizontal gene transfer from bacteria have a role?" Trends Genet. 20(7): 292-9.

Kaplan, M. M. and M. Flavin (1965). "Threonine biosynthesis. On the pathway in fungi and bacteria and the mechanism of the isomerization reaction." J. Biol. Chem. 240(10): 3928-33.

Karsten, W. E. and P. F. Cook (2002). "Detection of intermediates in reactions catalyzed by PLP-dependent enzymes: O-acetylserine sulfhydrylase and serine-glyoxalate aminotransferase." Methods Enzymol. 354: 223-237.

Kasting, R. and A. J. McGinnis (1958). "Use of glucose labeled with carbon-14 to determine the amino acids essential for an insect." Nature 182: 1380-1.

Kataoka, H., N. Sakiyama, et al. (1991). "Distribution and contents of free O-phosphoamino acids in animal tissues." J. Biochem. 109(4): 577-80.

Krishna, S. S., T. Zhou, et al. (2001). "Structural basis for the catalysis and substrate specificity of homoserine kinase." Biochemistry 40(36): 10810-8.

Laber, B., K. P. Gerbling, et al. (1994). "Mechanisms of interaction of Escherichia coli threonine synthase with substrates and inhibitors." Biochemistry 33(11): 3413-3423.

Laber, B., S. D. Lindell, et al. (1994). "Inactivation of Escherichia coli threonine synthase by DL-Z-2-amino-5-phosphono-3-pentenoic acid." Arch. Microbiol. 161(5): 400-403.

Laber, B., W. Maurer, et al. (1999). "Characterization of recombinant Arabidopsis thaliana threonine synthase." Eur. J. Biochem. 263(1): 212-221.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature 227(259): 680-685.

Malumbres, M., L. M. Mateos, et al. (1995). "Molecular cloning of the hom-thrC-thrB cluster from Bacillus sp. ULM1: expression of the thrC gene in Escherichia coli and corynebacteria, and evolutionary relationships of the threonine genes." Folia Microbiol. (Praha) 40(6): 595-606.

Malumbres, M., L. M. Mateos, et al. (1994). "Analysis and expression of the thrC gene of Brevibacterium lactofermentum and characterization of the encoded threonine synthase." Appl. Environ. Microbiol. 60(7): 2209-19.

Mas-Droux, C., V. Biou, et al. (2006). "Allosteric threonine synthase. Reorganization of the pyridoxal phosphate site upon asymmetric activation through S-adenosylmethionine binding to a novel site." J. Biol. Chem. 281(8): 5188-96.

Mehta, P. K. and P. Christen (2000). "The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Adv. Enzymol. 74: 129-184.

Mozzarelli, A., A. Peracchi, et al. (1989). "Microspectrophotometric studies on single crystals of the tryptophan synthase a2b2 complex demonstrate formation of enzyme-substrate intermediates." J. Biol. Chem. 264(27): 15774-15780.

O'Brien, P. J. and D. Herschlag (1999). "Catalytic promiscuity and the evolution of new enzymatic activities." Chem. Biol. 6(4): R91-R105.

Omi, R., M. Goto, et al. (2003). "Crystal structures of threonine synthase from Thermus thermophilus HB8: conformational change, substrate recognition, and mechanism." J. Biol. Chem. 278(46): 46035-45.

Payne, S. H. and W. F. Loomis (2006). "Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences." Eukaryot. Cell 5(2): 272-6.

Percudani, R. and A. Peracchi (2003). "A genomic overview of pyridoxal-phosphate-dependent enzymes." EMBO Rep. 4(9): 850-854.

Pils, B. and J. Schultz (2004). "Inactive enzyme-homologues find new function in regulatory processes." J. Mol. Biol. 340(3): 399-404.

Ramos, C. and I. L. Calderon (1994). "Biochemical evidence that the Saccharomyces cerevisiae THR4 gene encodes threonine synthetase." FEBS Lett. 351(3): 357-359.

Reeds, P. J. (2000). "Dispensable and indispensable amino acids for humans." J. Nutr. 130(7): 1835S-4180S.
Rost, B. (2002). "Enzyme function less conserved than anticipated." J. Mol. Biol. 318(2): 595-608.

Schildkraut, I. and S. Greer (1973). "Threonine synthetase-catalyzed conversion of phosphohomoserine to a-ketobutyrate in Bacillus subtilis." J. Bacteriol. 115(3): 777-785.

Skarstedt, M. T. and S. Greer (1973). "Threonine synthetase of Bacillus subtilis. The nature of an associated dehydratase activity." J. Biol. Chem. 248(3): 1032-1044.

Sterk, M. and H. Gehring (1991). "Spectroscopic characterization of true enzyme-substrate intermediates of aspartate aminotransferase trapped at subzero temperatures." Eur. J. Biochem. 201(3): 703-7.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res. 25(24): 4876-4882.

Tsai, C. H. and L. M. Henderson (1974). "Degradation of O-phosphohydroxylysine by rat liver. Purification of the phospho-lyase." J. Biol. Chem. 249(18): 5784-9.

Vigetti, D., L. Pollegioni, et al. (2002). "Property comparison of recombinant amphibian and mammalian allantoicases." FEBS Lett. 512(1-3): 323-8.

Zhou, T., M. Daugherty, et al. (2000). "Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily." Structure Fold. Des. 8(12): 1247-57.

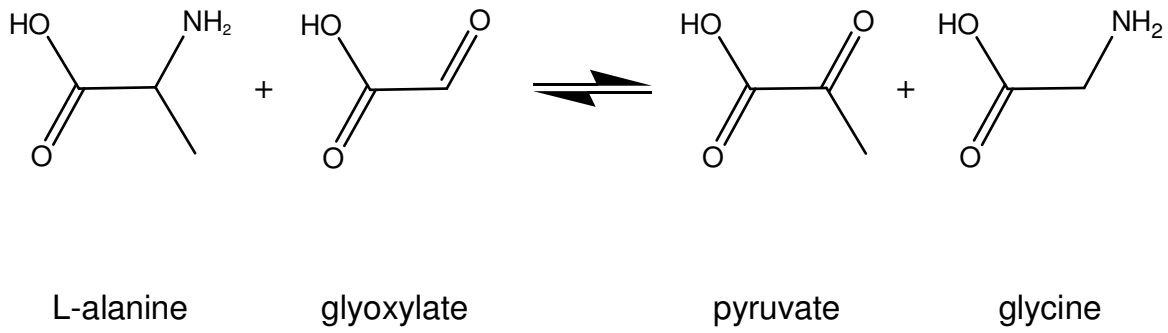
Chapter 3

Molecular cloning and initial characterization of two putative alanine-glyoxylate aminotransferase

Introduction

Alanine-glyoxylate aminotransferase (AGT; EC 2.6.1.44) is pyridoxal 5'-phosphate-dependent enzyme, distributed in a wide range of animal species. It employs the organic cofactor to transfer the amino group of L-alanine to glyoxylate, producing pyruvate and glycine (Rowsell, 1969; Noguchi, 1978; Rowsell, 1972; **Scheme 1**).

Scheme 1



Scheme 1. Reaction catalyzed by alanine-glyoxylate aminotransferase.

Two forms of AGT were identified in animal cells (Noguchi, 1978; Noguchi, 1987); one, designated AGT1, has an M*r* of 80KDa with 2 identical subunits, and is present both in the mitochondria and in the peroxisomes. AGT enzymes from both organelles show identical physical, enzymatic and immunologic properties. The AGT, designated AGT2, has an M*r* of 220KDa with 4 identical subunits. AGT1 is identical with serine-pyruvate aminotransferase (EC 2.6.1.51) and shows broad substrate specificity; in contrast, AGT2 is present only in the mitochondria, and is specific for L-alanine and glyoxylate (Noguchi, 1978; Takada, 1982).

The structure of human AGT1 has been solved (Zhang, 2003; Meyer, 2005), and the overall fold is similar to other enzymes of the fold-type I group (Grishin, 1995; Mehta & Christen, 2000; Christen & Mehta, 2001) the largest and the functionally most diverse family of PLP-dependent enzymes. Representative members of this group, in addition to alanine-glyoxylate aminotransferase, are 2,2-dialkyglycine decarboxylase, 4-aminobutyrate aminotransferase and ornithine aminotransferase (Mehta & Christen, 2000).

Clinical interest arose around alanine-glyoxylate aminotransferases because of their key role in glyoxylate detoxification. In fact, an AGT deficiency is the cause of Primary hyperoxaluria type 1 (PH1), a rare autosomal recessive disease characterized by progressive kidney failure due to renal deposition of calcium oxalate (Latta, 1990; Coulter-Mackie, 2004; Danpure, 2006). In AGT deficiency, glyoxylate is converted to oxalate, which forms insoluble calcium salts that accumulate in the kidney and other organs, culminating in renal failure and systemic oxalosis (**Figure 1**).

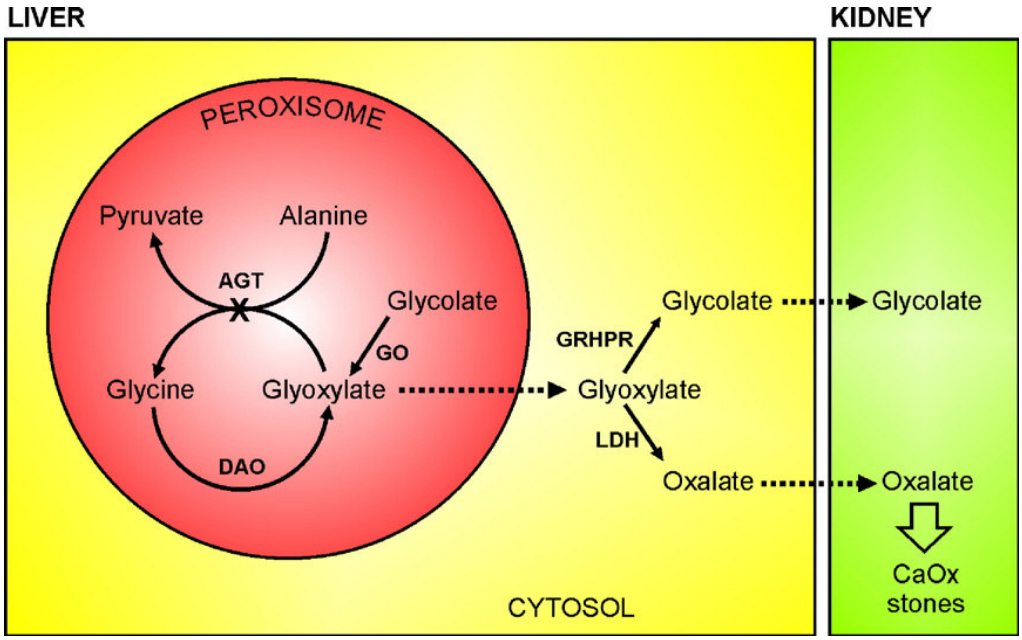


Figure 1. Metabolic consequences of peroxisomal AGT deficiency in PH1. AGT, alanine:glyoxylate aminotransferase; DAO, D-amino acid oxidase; GO, glycolate oxidase; GRHPR, glyoxylate/hydroxypyruvate reductase; LDH, lactate dehydrogenase. Solid arrows, metabolic conversions; dotted arrows, membrane or other transport.

In a recent bioinformatics study, a series of complete or near-complete genomes were scanned to obtain an inventory and a preliminary classification of genes that encode PLP-dependent enzymes (Percudani & Peracchi, 2003).

This analysis identified, in the genomes of several vertebrates (including man), two genes showing a high similarity to the second form of alanine-glyoxylate aminotransferase (these homologs are called *AGXT2L1* and *AGXT2L2*). Both these genes coded for cytosolic proteins, this observation raised the intriguing possibility that an unknown cytosolic form of AGT might be expressed. If this hypothesis were confirmed, new perspective would open for treatment of primary hyperoxaluria-1.

However, these could be yet another example of genes identified as coding for enzymes with a defined activity (in this case, an alanine-glyoxylate aminotransferase activity) only based on sequence criteria, which however possess an activity different from the predicted (Roost, 2002).

Moreover, it is reasonable to expect the discovery of PLP-dependent enzymes with activities that are not yet described or characterized, several examples support this possibility (Abe, 2006; Brushaber, 1998; Donini, 2006; Wolosker, 1999).

The present chapter describes the bioinformatic analysis of the two AGT2-like genes, as well as the cloning of cDNAs encoding the two homologs, and the successive initial characterization of the expressed and purified recombinant proteins.

Materials and Methods

Materials

DNA primers were purchased from MWG Biotech (Ebersberg, Germany). Rabbit muscle lactate dehydrogenase (LDH) was purchased from Fluka. All the potential amino group donors were from Fluka, or Sigma-Aldrich. All the amino group acceptors were from Sigma-Aldrich. PLP, PMP, NADH and $NADP^+$ were from Fluka. All other chemicals were of the highest purity available.

Bioinformatics

Initial identification of alanine-glyoxylate aminotransferase-like (AGXTL) genes in animal genomes was achieved through a search of PLP-dependent enzymes in genomic sequences using profiles Hidden Markov Models (HMM; Percudani & Peracchi, 2003).

Prediction of protein localization sites in cells was inferred using two different programs, PSORT II (http://psort.nibb.ac.jp/form2.html; Nakai & Horton, 1997) and pTarget (http://bioapps.rit.albany.edu/pTARGET/; Guda & Subramaniam, 2005). Physical and chemical parameters for each protein were calculated by means of ProtParam (http://psort.nibb.ac.jp/form2.html; Nakai & Horton, 1997) .

The tissue-specific expression profile of *AGXT2L1* and *AGXT2L2* sequences was inferred using the UniGene's EST ProfileViewer at the National Center for Biotechnology Information (NCBI). Homology searches were performed at the B6 DataBase (http://160.78.90.170/cgi-bin/bioinformatics/B6db/home.pl) using protein Blast.

Multiple sequence alignments were produced with ClustalX (Thompson, 1997). Phylogenetic analysis of alanine-glyoxylate aminotransferase-like was performed with the Boostrap Neighbor-joining method as implemented in the program ClustalX and tree-files in Phylip format were interpreted into visual trees by TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

Molecular cloning of AGXT2L1 and AGXT2L2 cDNAs

Clones of the complete cDNAs for both *AGXT2L1 and AGXT2L2* (respectively, IMAGE Consortium ID 4797767 and 5220434, cDNA from different tissues) were obtained from RZPD (Berlin, Germany).

Both coding sequences were amplified from these clones, however the protocols were different. *AGXT2L1* was amplified using the Fidelity Taq DNA polymerase (Stratagene) and adopting the following PCR conditions: 5 min of initial denaturation at 94°C, 3 cycles of 30 s denaturation at 94°C and 30 s annealing at 46°C, followed by 120 s polymerization at 72°C, afterwards 25 cycles of 30 s denaturation at 95°C and 30 s annealing at 56°C, followed by 120 s polymerization at 72°C.

AGXT2L2 was also amplified using the Fidelity Taq DNA polymerase, but the adopted PCR conditions were different: 3 min of initial denaturation at 94°C, 25 cycles of 30 s denaturation at 94°C and 30 s annealing at 57°C, followed by 120 s polymerization at 72°C.

The gene-specific primers were:

AGXT2L1_cpo^{+}	atCGGTCCGatgtgcgagctgtacagtaagc
AGXT2L1_cpo^{-}	atCGGACCGatcttgctttaaaatgcaaatcag
AGXT2L2_cpo^{+}	atCGGTCCGatggccgcagaccagcgc
AGXT2L2_cpo^{-}	atCGGACCGtaggcagagcagggctggc

Underlined residues correspond to the start codon. The 5' region of all the primers contained sequences (uppercase letters) recognized by the restriction enzyme *Cpo*I, so that the amplification product contained restriction sites near to both ends.

Amplicons were directly cloned into the vector pBlueScript II KS+ (Stratagene) treated with *Sma*I (Bolchi, 2005). The cloned inserts were transformed into XL1B *Escherichia coli* cells (Stratagene) and sequence-verified. Then amplicons were

treated with *Cpo*I and subcloned into the expression vector pET28-H6-*Cpo* (digested with *Cpo*I, too).

This plasmid is a derivative of pET28 (Novagen, Madison, WI) modified to present a single *Cpo*I restriction site in the cloning region, downstream to a sequence encoding a hexahistidine tag. A schematic description of the molecular cloning strategy is shown in **Figure 2**.

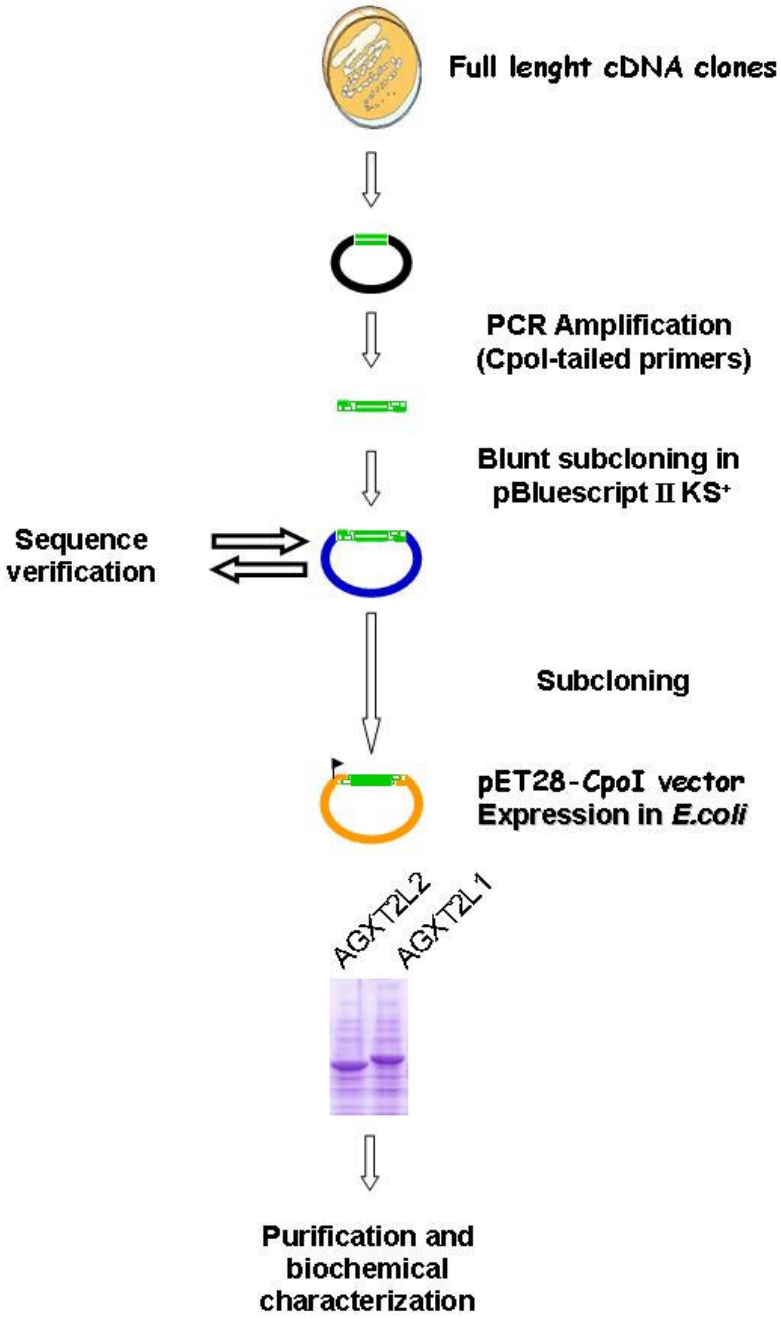


Figure 2. Overview of the general procedure adopted for the molecular cloning and recombinant expression of AGXT2L1 and AGXT2L2.

In vitro site-directed mutagenesis of the AGXT2L2 clone

The mutated AGXT2L2 was corrected using the QuikChange™ site-direct mutagenesis method (Stratagene). Two oligonucleotide primers containing the desired mutation were extended during PCR, using the Pfu Ultra DNA polymerase (Stratagene) and adopting the following conditions: 5 min of initial denaturation at 94°C, 12 cycles of 30 s denaturation at 95°C and 60 s annealing at 55°C, followed by 80 s polymerization at 72°C.

The primers were:

AGXT2L2_Anti	5'-ctgcgtggccgcaaACccagcctgtggcg-3'
AGXT2L2_Sense	5'-cgccacaggctggGTtgcggccacgcag-3'

Each primer is designed as recommended by the kit manufacturer. In the sequences, uppercase letters represent the desired mutation. Then, the corrected amplicon was transformed into XL1B *Escherichia coli* cells (Stratagene) and the identity of the construct was confirmed by DNA sequencing.

Heterologous gene expression and protein purification

The cloned inserts were then transformed into BL21-CodonPlus®-RIL *Escherichia coli* cells (BL21 codon$^+$; Novagene), BL21 Star™ *Escherichia coli* cells (BL21 Star, Novagene) and Origami *Escherichia coli* cells (Novagene).

Except for Origami, in the other two strains of *E. coli* the recombinant protein was expressed efficiently. However, BL21 Star was adopted as the expression system, because of the slightly higher fraction of soluble protein obtained.

For AGXT2L1 a subculture was used to inoculate 1 liter of Lauria-Bertani (LB) broth containing 50 $\mu g \bullet ml^{-1}$ kanamycin (Sigma). The cells were grown at 37°C in vigorous aeration and agitation. After ~ 3 h, when the cell density corresponded to an absorbance of ~ 0.6 at 600 nm, isopropyl β-D-thiogalactoside (1 mM) was added.

After aerobic culturing for 3 gg at 12°C, cell paste was obtained by centrifugation and stored at -80°C. The same procedure was employed to AGXT2L2 expression, but in this case a minimal medium (M9) was used for the cellular growth.

After cell sonication, the soluble lysate fraction was loaded onto a metal-affinity resin (Talon; Clontech) equilibrated in Buffer A (5% glycerol/ 300 mM NaCl/ 50 mM phosphate-Na buffer pH 7.0) supplemented with 40 μM PLP, 5 mM β-mercaptoethanol and protease inhibitors (1 μM leupeptine, 1 μM pepstatine, 0.5 mM benzamidine, 0.5 mM phenyl methyl sulfonyl fluoride). The recombinants AGXT2L1 and AGXT2L2, carrying an N-terminal hexahistidine tag, were purified following the manufacturer's instructions.

The composition and purity of each protein fraction was assessed by gel electrophoresis on SDS-polyacrylamide (11%) gels (Laemmli, 1970); fractions with a purity > 90% were pooled, dialyzed against buffer A supplemented with 4 μM PLP and 1 mM dithiothreitol, and stored at -80 °C.

The concentration of each purified recombinant protein was estimated by the Bradford assay (Bradford, 1976) with bovine serum albumin as a standard, and confirmed based on the calculated extinction coefficient at 280 nm (Gill & von Hippel, 1989).

Spectrophotometric measurements

Solutions containing recombinant protein (6 μM final) in 40 mM bis-tris propane (BTP) buffer (pH 8), were supplemented with various potential substrates (10 mM, unless specified otherwise). Spectral changes were monitored using a Cary 400 spectrophotometer (Varian). Cuvette holders were thermostated at 25 °C. Time-dependent spectral changes (time-courses) and titration data were analyzed by nonlinear least-squares fitting to the appropriate kinetic or thermodynamic equation using Sigma Plot (SPSS Inc.).

Enzymatic assays

The cysteine aminotransferase assay (200 μl final volume) contained 40 mM BTP buffer (pH 8), 1 mM of α-keto acid, 20 mM of L-Cys, 0.25 mM NADH, 1.4 units/ml lactate dehydrogenase, while the enzyme was 6 μM. Reactions were conducted at 37°C, and the rate of β-mercaptopyruvate production was measured following the disappearance of NADH at 340 nm.

The possible disappearance of L-cysteine and the formation of a new amino acid was tested by incubating L-Cys in the same reaction conditions of the cysteine

amino transferase assay and then running the reaction mixture on a silica TLC plate, side-by-side with amino acid standards (L-Cys and the L-amino acid corresponding to the α-keto acid tested). Amino acids were visualized with fuorescamine (Meola and Brown, 1974).

In order to assay the cysteine lyase activity, L-cysteine, and D-cysteine, were used as substrates and the rates of pyruvate production were measured spectrophotometrically by a coupled assay, in which the enzyme (6 μM) and substrate (10 mM) were incubated in the presence of lactate dehydrogenase and NADH. The reaction mixture (200 μl final volume) contained 40 mM BTP buffer (pH 8), 0.25 mM NADH, 2 units/ml lactate dehydrogenase. The reactions were conducted at 37°C, and the disappearance of NADH was monitored at 340 nm.

Thiazolidine formation assay

In order to detect the formation of thiazolidine (TZA), a general and simple method for the conjugation of L-cysteine and PLP was developed. This method is based on the specific reaction of the L-cysteine with the aldehyde group from cofactor under the same conditions used to evaluated the affinity of each enzyme with different ligands.

Solutions containing PLP (250 μM final) in 40 mM BTP buffer (pH 8), were supplemented with L-Cys (5 mM). Spectral changes were monitored using a Cary 400 spectrophotometer (Varian). Cuvette holders were thermostated at 30 °C. Spectral changes data were analyzed by nonlinear least-squares fitting to the appropriate kinetic or thermodynamic equation using Sigma Plot (SPSS Inc.).

Results

Identification of AGXT2L1 and AGXT2L2 genes

To help the identification and functional assignment of PLP-dependent enzymes, our group has created a database of PLP-dependent enzymatic activities and sequences (Percudani & Peracchi, 2003). While performing an inventory of PLP-dependent enzymes in different organisms it was noticed that several invertebrates and vertebrate genomes contained genes encoding proteins similar to the mitochondrial alanine-glyoxylate aminotransferase (AGT2), despite the fact that two other alanine-glyoxylate aminotransferase (AGXT1 and AGXT2) are known to occur in vertebrates.

Notably, in vertebrates these alanine-glyoxylate aminotransferase-like (AGXTL) sequences are divided into two enzyme subtypes, AGXT2L1 and AGXT2L2, which generally were co-present in the same organism. A list of genes, some of which are annotated as AGXTL in public databases, is presented in **Table 1**.

The human *AGXT2L1* gene encodes a 499-amino acid protein with a 36% identity and 55% similarity to the human AGT2. While, the human *AGXT2L2* gene encodes a 450-amino acid protein with 36% identity and 58% similarity to AGT2. The most obvious difference between the two genes lies in a 40-aa C-terminal tail which occurs in the AGXT2L1 lineage and is absent in the *AGXT2L2* gene products.

Moreover, multiple sequence alignment of the AGXT2L1 and AGXT2L2 revealed a ~ 30% identity with enzymes of the AT II subfamily in the fold-type I family (Mehta & Christen, 2000; data not show).

In silico characterization of AGXT2L1 and AGXT2L2 has predicted molecular mass of 58 kDa and 52 kDa, respectively. According to several programs predicting the intracellular localization of proteins in the cell, both the enzymes should be cytoplasmic.

Examination of the ESTs databases indicated that mRNAs for both genes are present in tissues from their respective organisms, confirming that the genes are expressed. Based on the high frequency with which transcripts are found in public EST libraries, expression of *AGXT2L1* and *AGXT2L2* appears to occur at high levels in humans. *AGXT2L1* gene transcripts are observed predominantly in the brain, kidney and liver. *AGXT2L2* transcripts are also detected in the brain, kidney and liver, however this gene is expressed in almost all the tissues.

Table 1
List of metazoan genes encoding AGT2 or AGT2-like proteins

Group	Organism	Abbreviation	GenBank/Ensembl Accession #	Predicted Protein Size (aa)[a]
AGT2	*Homo sapiens*	Hsa	NP_114106	514
	Rattus norvegicus	Rno	BAA07281	512
AGXT2L	*Tribolium castaneum*	Tca	XP_969816	472
	Caenohrabditis briggsae	Cbr	XP_001671708	467
	Caenohrabditis elegans	Cel	NP_001023346	467
	Ciona intestinalis	Cin	AK114044	427
	Glossina morsitans morsitans	Gmm	ABG77981	494
	Aedes aegypti	Aae	XP_001656046	490
	Apis mellifera	Ame	XP_392348	482
	Strongylocentrotus purpuratus	Spu	XP_001187418	531
	Anopheles gambiae	Aga	XP_310808	462
	Drosophila melanogaster	Dme	NP_648665	494
AGXT2L1	*Homo sapiens*	Hsa1	Q8TBG4	499
	Macaca fascicularis	Mfa1	XP_001087348	499
	Bos taurus	Bta1	NP_001015605	497
	Mus musculus	Mmu1	Q8BWU8	499
	Gallus gallus	Gga1	XP_426301	505
	Danio rerio	Dre1	Q7SY54	492
	Xenopus laevis	Xla1	Q6DEB1	509
	Rattus norvegicus	Rno1	XP_001076965	472
AGXT2L2	*Homo sapiens*	Hsa2	Q8IUZ5	450
	Macaca fascicularis	Mfa2	XP_001095786	450
	Bos taurus	Bta2	XP_617398	450
	Mus musculus	Mmu2	Q8R1K4	467
	Gallus gallus	Gga2	XP_414688	533
	Canis familiaris	Cfa2	XP_538569	450
	Danio rerio	Dre2	NP_001032646	447
	Tetraodon nigroviridis	Tni2	CAG01207	434

Table 1. The group, organism source, abbreviation, databank accession number, protein length, are shown for all the genes encoding alanine-glyoxylate aminotransferase-like proteins, or validated AGT2. [a]When multiple splice variants of the gene product are included in the databases, the size refers to the longest predicted polypeptide.

Comparison and phylogeny of AGXT2L gene products

Preliminary phylogenetic of all the 26 amino acid sequences for AGT-like proteins illustrates the differences between vertebrates and invertebrates. In fact, the genomes of invertebrates such as *C. elegans* and insects contain a single AGT2-like gene. While, AGXT2L1 and AGXT2L2 seem derived from duplication of a single ancestor during metazoan evolution (**Figure 3**).

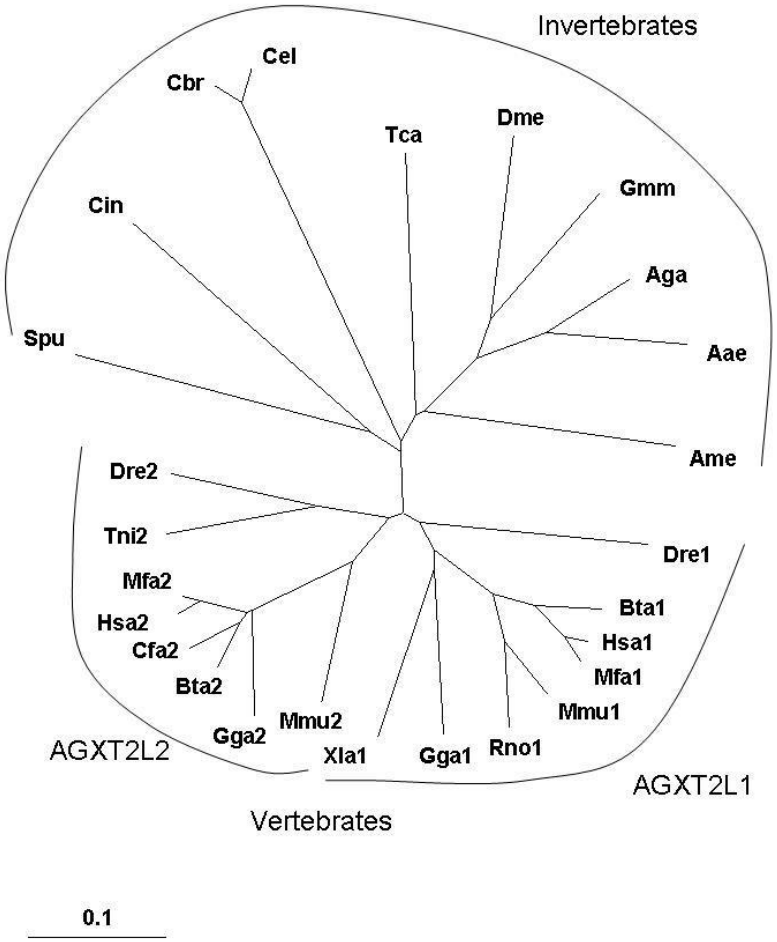


Figure 3. Phylogenetic tree of AGXTL proteins. The unrooted Neighbor-joining phylogenetic tree was constructed with the program ClustalX and displayed with the TreeView program. Sequences are indicated with abbreviations in **Table 1**.

Multiple sequence alignment of AGT-like proteins with validated AGTs, confirmed significant homology, the catalytic lysine are conserved, too (**Figure 4**; Lee, 1995). Of the two vertebrate gene products, the AGXT2L2 proteins are slightly more divergent from the presumed ancestor.

This may be appreciated especially by looking at a few diagnostic positions: in particular, at the position corresponding to C106 in the human AGXT2L1, all AGXT2L1 and invertebrate AGXT2L contain a cysteine residue, whereas the AGXT2L2 proteins contain a phenylalanine; similarity, at position 218, all AGXT2L1 and most invertebrate AGXT2L proteins contain cysteine, whereas the AGXT2L2 proteins contain Val; at position 211, AGXT2L1 and invertebrates AGXTL contain Ile, but AGXT2L2 proteins contain Phe (**Figure 4**).

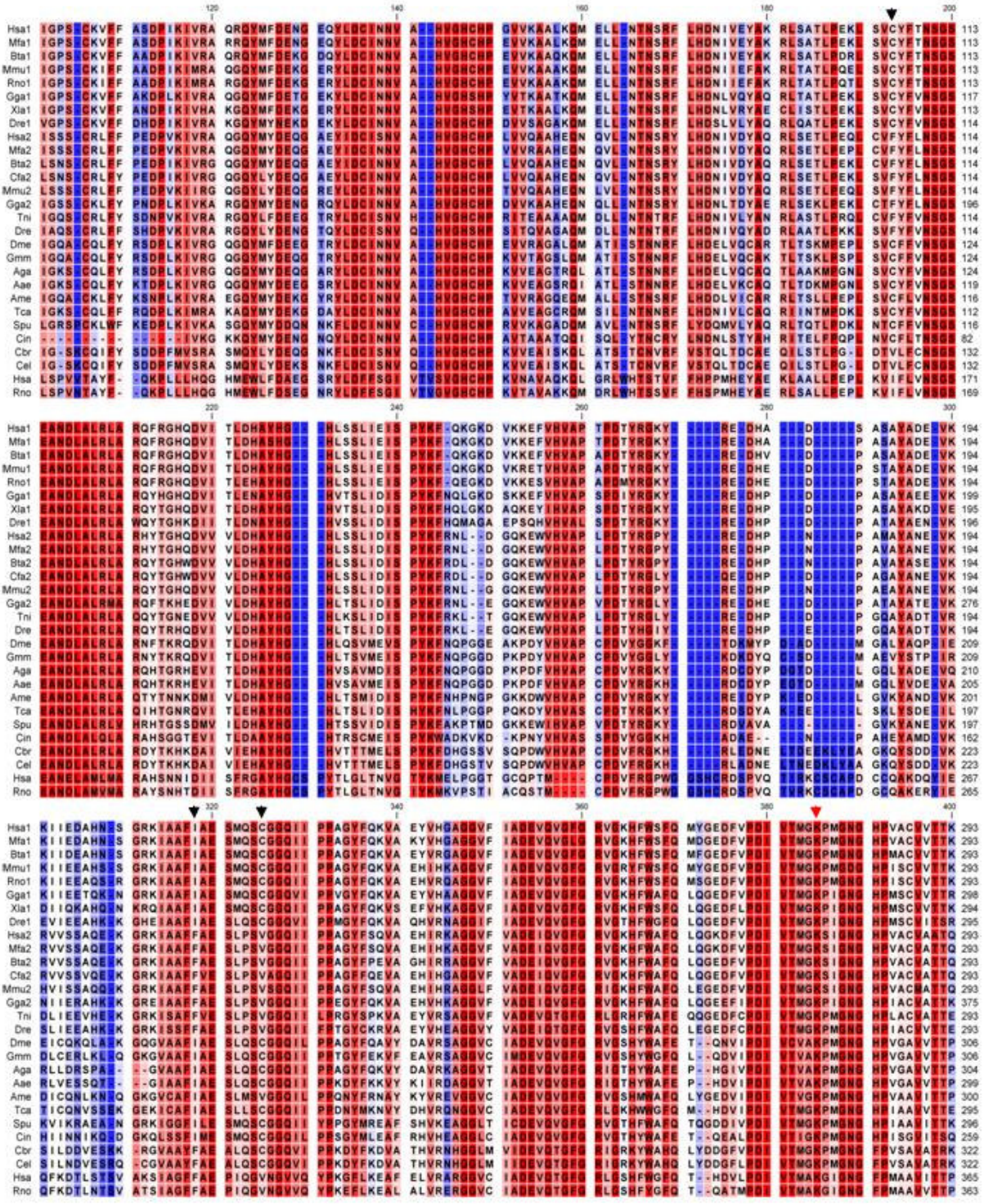


Figure 4. Comparison of the amino acid sequence around the PLP-binding sites of AGT from *Homo sapiens* (NP_114106), and *Rattus norvegicus* (BAA07281) with AGT2-like proteins. A color-scale represents the conservation: residues that are absolutely conserved are red, instead residues that are not conserved are blue. The black arrows signal those residues that are mutated between AGXT2L1, and AGXT2L2. The red arrow marks the catalytic lysine residue that binds the PLP cofactor through a Schiff base linkage.

Unfortunately there are not crystal structures of AGT2, showing the arrangement of residue involved in the active site. In order to understand how the active site could be organized, the amino acid sequences of AGXT2L1 and AGXT2L2 were compared with other enzymes classified in the AT II subfamily, such that of AGT2 (Lee, 1995), which was characterized their structure, and the role played in the active site (**Table 2**).

Notably, AGXT2L1 and AGXT2L2 were significantly homologous to *P. cepacia* 2,2-dialkylglycine decarboxylase (Fogle, 2005; Keller, 1990; EC 4.1.1.64), *H. sapiens* ornithine aminotransferase (Shen, 1998; EC 2.6.1.13) and *S. scrofa* 4-aminobutyrate aminotransferase (Storici, 1999; Storici, 2004; EC 4.6.1.19). In fact, the active site sequences were well conserved among these validated enzymes, mainly in the PLP-binding region (**Figure 5**; **Table 2**).

Table 2
Comparisons of the active site residues involved in the PLP-binding

Role in the active site	AGXT2L1	AGXT2L2	DGD	OAT	GABA-AT
Catalytic Lysine	K_{278}	K_{278}	K_{270}	K_{292}	K_{329}
Binds N1 PLP	D_{246}	D_{246}	D_{241}	D_{263}	D_{298}
Binds O3 PLP	Q_{249}	Q_{249}	Q_{244}	Q_{266}	Q_{301}
Supports the pyridine ring	V_{248} Y_{139}	I_{248} Y_{140}	A_{243} W_{1136}	I_{265} F_{177}	V_{300} F_{189}
Binds the oxygens of the Pi group	S_{113} G_{112} T_{309}	S_{114} G_{113} T_{309}	A_{110} G_{109} T_{302}	V_{143} G_{142} T_{322}	S_{137} G_{136} T_{353}

Table 2. Topographical alignment of the active site residues among various PLP-dependent enzymes of the AT II subfamily and both the recombinant enzymes. All the structures of the enzymes considered herein, except AGXT2L1 and AGXT2L2, have been characterized. Legend: *P. cepacia* 2,2-dialkylglycine decarboxylase (DGD), *H. sapiens* ornithine aminotransferase (OAT), *S. scrofa* 4-aminobutyrate aminotransferase (GABA-AT).

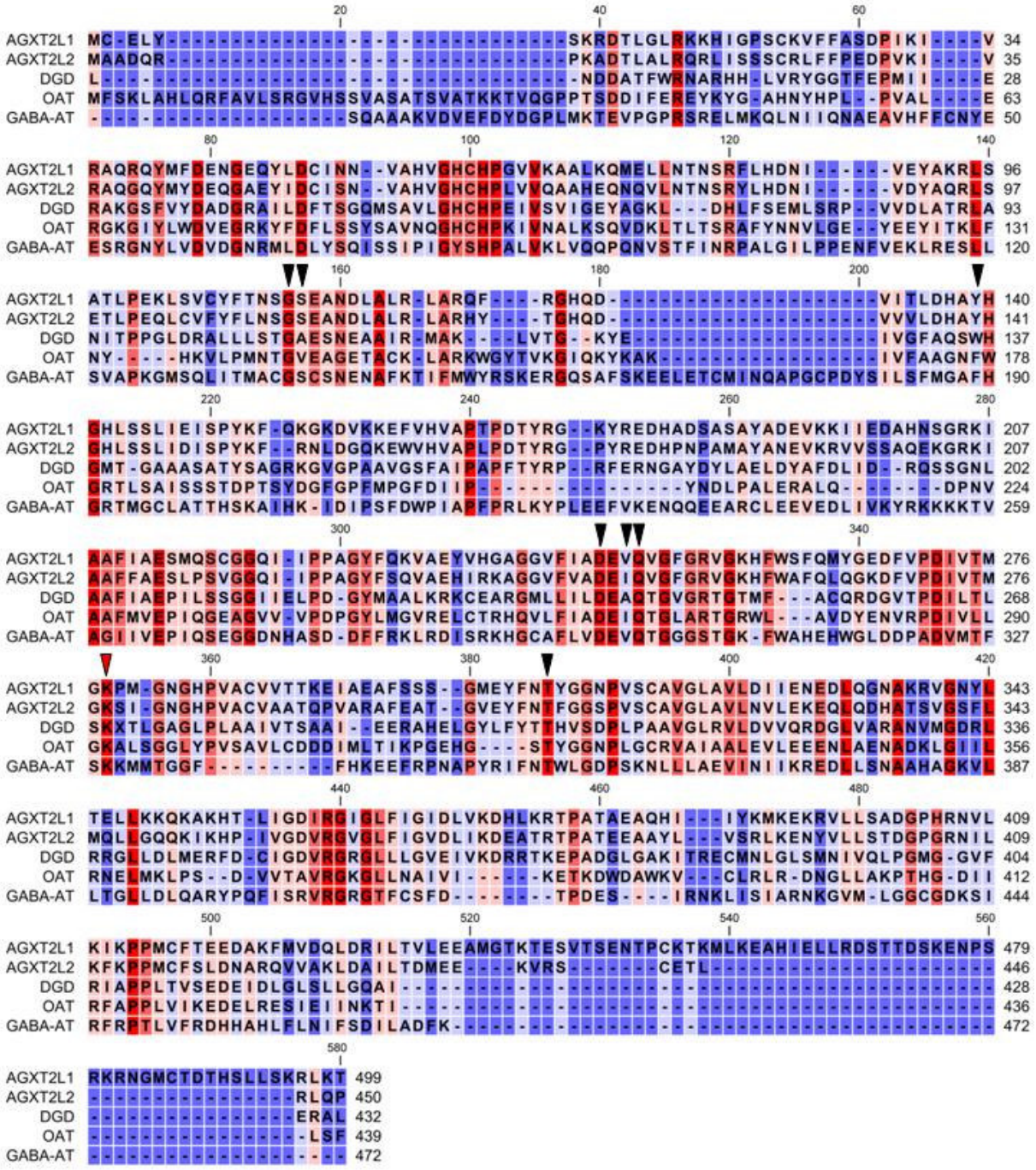


Figure 5. Alignment of the amino acid sequences of AGXT2L1 and AGX2L2, with those of *P. cepacia* 2,2-dialkylglycine decarboxylase (DGD), *H. sapiens* ornithine aminotransferase (OAT), *S. scrofa* 4-aminobutyrate aminotransferase (GABA-AT). A color-scale represents the conservation: identical residues are red, instead residues that are not conserved are blue. The black arrows signal those residues that bind the PLP cofactor. The red arrow marks the catalytic lysine residue that binds the PLP cofactor through a Schiff base linkage.

While the closest validated homologs of AGXT2L sequences are indeed metazoan AGT2 enzymes, there are other proteins of unknown function that show an even higher similarity to AGXT2L. These are (1) a series of peroxisomal plant enzymes that do not catalyze the transamination of glyoxylate and for whom the metabolic significance remains obscure (Liepman, 2003); (2) a series of bacterial enzymes (for example NP_437638 from *Sinorhizobium*) that are normally expressed as fusion proteins with an additional domain. This additional domain codes for a homoserine kynase type II, but this function is only predicted; in fact, this definition apparently comes from a previous bioinformatic study (Gophna, 2005). The domain is also annotated as belonging to a phosphotransferase enzyme family that includes bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides.

Cloning, expression and purification of recombinant AGXT2L1 and AGXT2L2

The full-length human genes nucleotide sequences obtained from the IMAGE consortium (http://image.llnl.gov) were utilized to amplify the cDNAs fragment by PCR. Since *Escherichia coli* is by far the most widely employed host, and was used for example for the expression of wild type AGT1 (Lumb, 2000; Coulter-Mackie, 2004), and since post-translational modifications of the products are presumably not essential, this organism was used for the expression of recombinant AGXT2L1 and AGXT2L2 (see Methods).

Attempts to subclone *AGXT2L1* in an *E. coli* expression vector were successful: several positive colonies were obtained, and the cloned insert, after sequence verification, did not contain frameshifts or deletions. However, when subcloning *AGXT2L2* in the same *E. coli* expression vector, only one positive colony was observed; moreover, after sequence verification, the cloned insert contained a missense mutation, such that the mutated codon (acc → gac) coded for Asp instead of Thr. By employing the QuikChange™ site-direct mutagenesis method (Stratagene), the original gene sequence was restored.

The expression of both recombinant proteins in *E. coli* BL21 Star, and their purification by cobalt-resin affinity chromatography are described in the Materials and Methods. A great yield of soluble AGXT2L1 (~50% of the total recombinant protein) was obtained; in contrast, overexpression of AGXT2L2 protein was accomplished, but soluble protein could only be obtained in low yields (~10% of the total recombinant protein).

Recombinant AGXT2L1 was 521 residues long (including the N-terminal hexahistidine tag) with a calculated molecular mass of 58 kDa. While AGXT2L2 was 472 residues long (the N-terminal hexahistidine tag is included) with a calculated molecular mass of 52 kDa. Purified AGXT2L1 and AGXT2L2 were homogenous by SDS-PAGE (data not show), and exhibited the characteristic PLP absorption maximum at 408 nm for AGXT2L1 (**Figure 6**) and at 412 nm for AGXT2L2 (**Figure 7**).

Interaction of AGXT2L1 and AGXT2L2 with potential substrates

AGT, like many other aminotransferases, operates by the same basic mechanism. Initially, L-alanine reacts with the pyridoxal form of the enzyme (E-PLP) to yield the keto acid (pyruvate) and the pyridoxamine form of the enzyme (E-PMP). In the second half-transamination, glyoxylate reacts with E-PMP to reform E-PLP and glycine (**Scheme 1**).

Usually, for aminotransferases, the pyridoxal form and pyridoxamine form of the enzyme show different absorption peaks, with a maxima at 410-420 nm (for E-PLP) and at ~ 330-340 nm (for E-PMP). Human alanine-glyoxylate aminotransferase 1 has been characterized by means of these spectroscopic properties (Cellini, 2007).

Thus, AGXT2L1 and AGXT2L2, could be preliminarily characterized by monitoring the binding of substrates, substrate analogs and inhibitors, as well as the formation of catalytic intermediates by UV-absorption spectroscopy (e.g., Laber, 1994; Mozzarelli, 1989; Sterk & Gehring, 1991; Karsten & Cook, 2002).

First, the interaction of AGXT2L1 with L-Ala, or Gly (the physiological substrates of AGT) was tested spectroscopically. However, L-Ala and Gly did not induced absorption change spectra (data not shown), hinting that the enzyme may not be binding these amino acids. Analogous results were obtained with most of the other standard amino acids tested as ligands.

In contrast, significant spectral changes were observed with L-Cys and L-Glu. Noteworthy, upon addition, L-cysteine induced a bigger change in absorbance, than L-Glu, suggesting an higher affinity for this ligand (**Figure 6A**).

Mixing of AGXT2L1 with L-Cys, induces in the immediate appearance of a maximum band at 330 nm, and the disappearance of maximum band at 410 nm. This kind of change in absorbance takes place during the first half-transamination; in fact, the peak at 408 nm represents the E-PLP, while the peak at 330 nm resembles the typical absorption peak for E-PMP (**Figure 6A**).

In contrast, addition of L-Glu at the same concentration did not induce an immediate change in absorption, but several minutes were necessary to detect spectral changes attributable to the first half-transamination (**Figure 6A**).

Based on to the above results, compounds that show similarity to L-cysteine were tested spectroscopically. Only DL-homocysteate and cystathionine did not induce absorption changes (data not shown), while the addition of D-cysteine, DL-homocysteine, or cysteamine resulted in absorption changes compatible with the appearance of the species E-PMP, and the disappearance of E-PLP (**Figure 6B**).

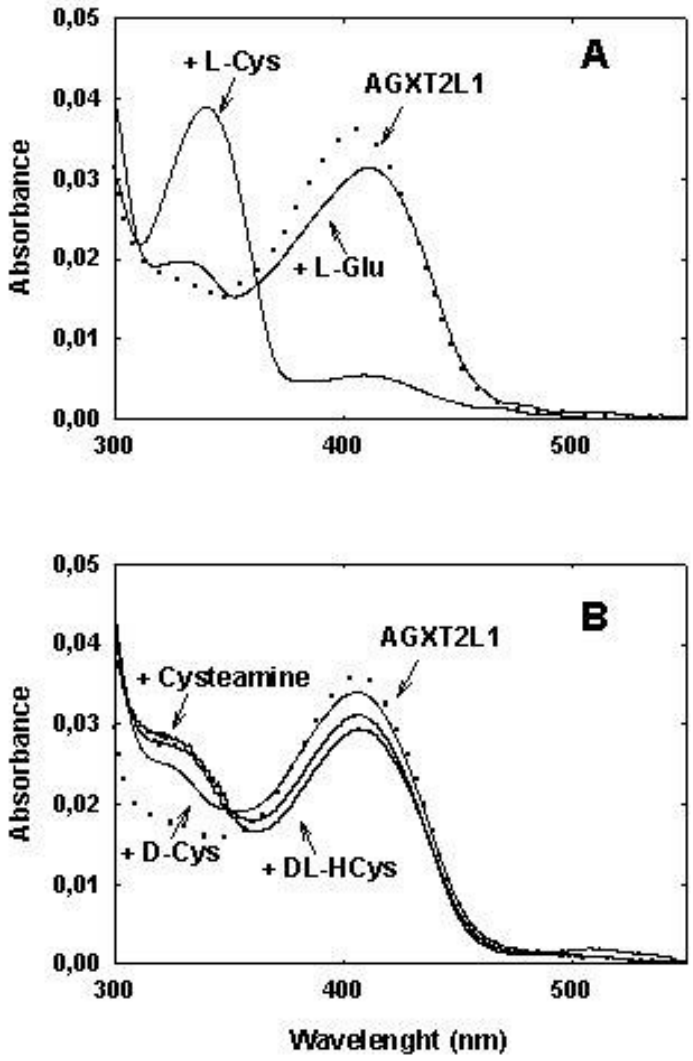


Figure 6. Absorption spectra of AGXT2L1 in the presence and absence of amino acid ligands. The protein (6 μM) was incubated in 40 mM BTP, pH 8, 37°C. (A) Absorption spectra were recorded in the absence of ligands (dotted line) or in the presence of 10 mM L-Cys or 10 mM L-Glu (after 10'). (B) Absorption spectral change of AGXT2L1 (dotted line) in the presence of L-Cys analogous (ligands final concentrations was 10 mM). The ligands tested were D-cysteine (D-Cys), Cysteamine, DL-homocysteine (DL-HCys).

Also for AGXT2L2 the interaction with L-Ala, or Gly (the physiological substrates of AGT) were tested spectroscopically. This enzyme, too, did not show affinity for these ligands. Moreover, as with AGXT2L1, most of the other standard amino acids, except L-Cys, failed to give appreciable absorption changes (**Table 3**).

However, this enzyme did not show a response identical to AGXT2L1. In fact, only binding of L-Cys was detected, instead addition of L-Glu did not induced absorption changes (**Figure 7A**). In order to compare the substrate specificity between these two PLP-dependent enzymes, compounds that show similarity to the substrate were tested spectroscopically in presence of AGXT2L2 (**Figure 7B**).

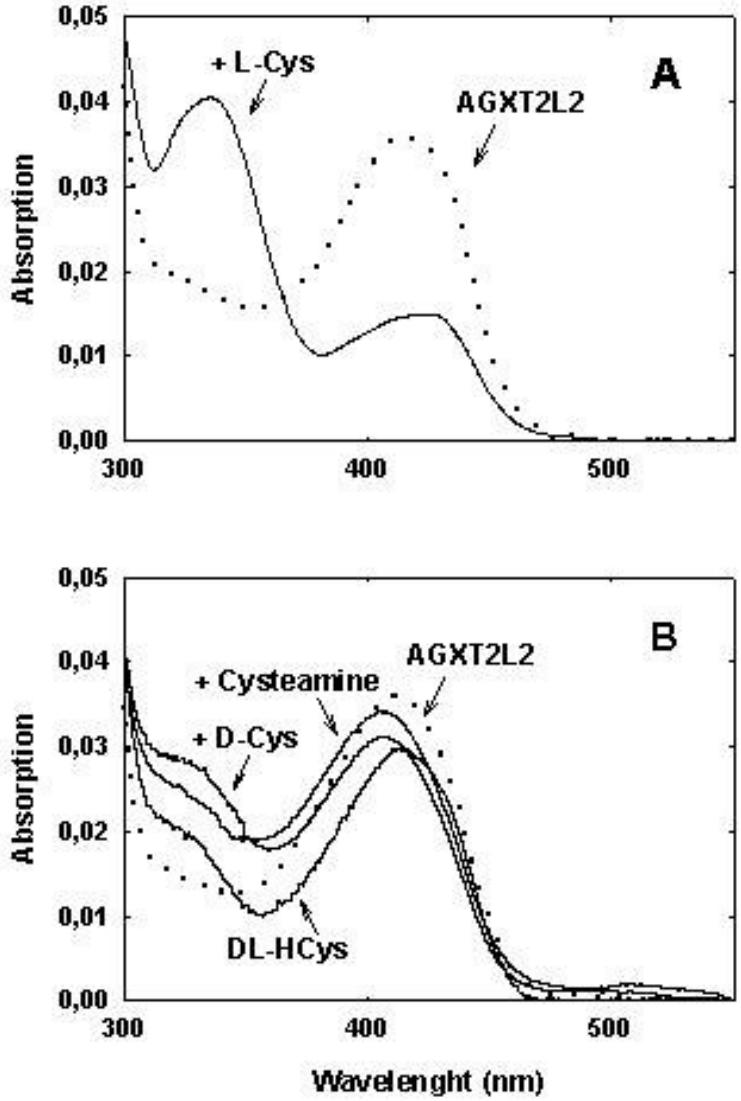


Figure 7. Absorption spectra of the purified recombinant AGXT2L2 in the presence and absence of amino acid ligands. Enzyme in native form (6 μM) was incubated in 40 mM BTP, pH 8, 37°C. (A) Absorption spectra were recorded in the absence of ligands (dotted line) or in the presence of 10 mM L-Cys or 10 mM L-Glu. (B) Absorption spectral change of AGXT2L2 (dotted line) in the presence of L-Cys analogous (ligands final concentrations was 10 mM). The compound tested were D-cysteine (D-Cys), Cysteamine, DL-homocysteine (DL-HCys).

Similar to AGXT2L1, the absorption spectrum of AGXT2L2 was changed in presence of the same ligands, however some differences were observed (**Table 3**); addition of D-cysteine induced formation of an higher peak at 330nm, suggesting better affinity of this enzyme for this ligand. In contrast, with cysteamine, AGXT2L2 seems to show lower affinity. Only addition of DL-homocysteine gave similar results compared to the AGXT2L1 spectra (**Figure 7B**).

Table 3
List of ligands tested

Ligand	AGXT2L1	AGXT2L2
glycine	-	-
L-alanine	-	-
L-methionine	-	-
L-valine	-	-
L-isoleucine	-	-
L-leucine	-	-
L-lysine	-	-
L-arginine	-	-
L-threonine	-	-
L-serine	-	-
L-tryptophan	-	-
L-phenylalanine	-	-
L-tyrosine	-	-
L-glutamate	+/-	-
L-glutamine	-	-
L-aspartate	-	-
L-asparagine	-	-
L-histidine	-	-
L-cysteine	++++	++++
D-cysteine	++	+++
DL-homocysteine	*+++*	+++
cysteamine	*+++*	*++*
L-ornithine	-	-
L-kynurenine	-	-
γ-aminobutyric acid	-	-
DL-homocysteate	-	-
cystathionine	-	-

Table 3. Ligands specificity of AGXT2L1 and AGXT2L2. relative specificity for different substrates is valuated measuring the change in ratio E-PMP/E-PLP. Spectra were perform at the same condition for all the compound (see Materials and Method).

Cysteine-oxoglutarate aminotransferase activity assay of AGXT2L1

Based on its interactions with L-Cys and L-Glu, AGXT2L1 could be hypothesized to be a cysteine-oxoglutarate aminotransferase (CGT); notably, examples of this activity are documented in the literature (Ip, 1977; Akagi, 1982).

CGT activity was assayed for AGXT2L1. The activity was measured as described in the Materials and Methods. However, negligible, or inexistent, β-mercaptopyruvate production was detected (data not show).

Also, when AGXT2L1 was incubated with cysteine and oxoglutarate (under the same conditions of the coupled assay) for a few hours and the reaction products were analyzed by TLC followed by fluorescamine coloration, no evidence was found for the generation of L-Glu and disappearance of L-Cys (data not show).

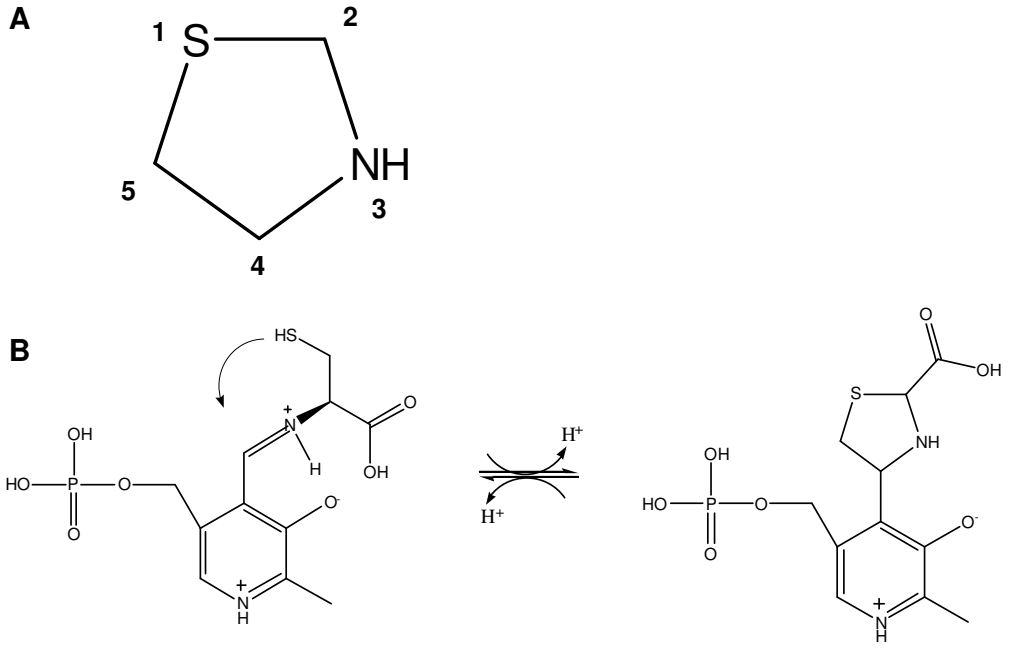
The spectral changes induced by L-Cys are due to formation of a thiazolidine adduct

Thiazolidine (TZA) is a 5-membered saturated ring with an amine group and a sulfide group in 1 and 3 positions (**Scheme 2**; Canle, 1996), that can form upon reaction of PLP with various amino thiols (Ponticelli, 1983).

In principle, the spectrum of the thiazolidine formed by L-Cys with PLP should resemble the spectrum of PMP. If the cromophoric species shown for example in **Figure 6A** were a thiazolidine adduct, this could explane the absence of activity as CGT for AGXT2L1 (**Scheme 2**).

To test this hypothesis the formation of a TZA was assayed, in presence of L-cysteine and PLP, by a spectrophotometric method as described (see in Matherials and Methods). When the amino acid was added the PLP spectrum showed the appearance of a band with maximum at 330 nm, undistinguishable from the peak of PMP. This datum supports the hypothesis that both the recombinant enzymes can bind L-Cys, in an unproductive way (data not show).

Scheme 2



Scheme 2. Eventual unproductive bind with L-Cys**.** (A) Thiazolidine ring. (B) Hypothetical mechanism for the formation of the TZA-complex after binding with PLP.

AGXT2L1 and AGXT2L2 catalyzes slow desulfhydrase reactions with L-Cys and D-Cys

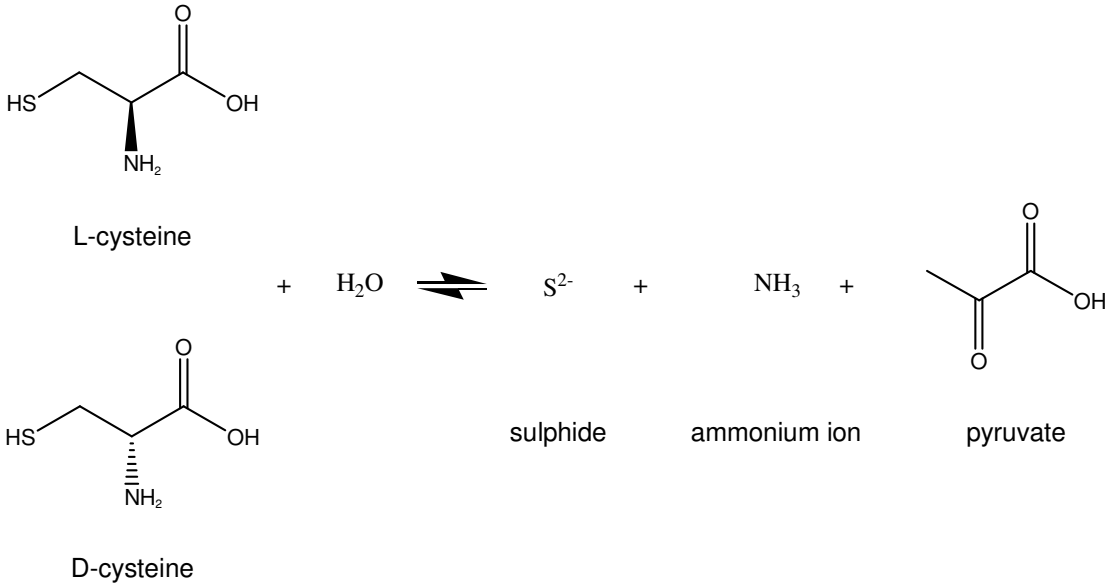
Even though AGXT2L1 and AGXT2L2 are not cysteine transaminases, their high apparent specificity for L-Cys, and similar substrates, led us to test their activity as desulfhydrases.

The purified enzymes were probed for this activity through the measurement of the pyruvate formed from L-cysteine, or D-cysteine, by using a spectrophotometric method with lactate dehydrogenase and NADH as described (see Matherials and Methods).

Results were different between the two enzymes. L-cysteine, and D-cysteine, for AGXT2L2 were substrates. In contrast, reaction with AGXT2L1 showed a very slow formation of pyruvate (and presumably sulphide and ammonium ion), from both the tested substrates, suggesting that the enzyme can catalyze the cleavage of the C-S bond (**Scheme 3**).

When the kinetics were measured (described in Matherials and Methods) the calculated specific activity was 0.0114 μmol min^{-1} mg^{-1} for L-cysteine as substrate, and 0.0075 μmol min^{-1} mg^{-1} with D-cysteine.

Scheme 3



Scheme 3. Reaction catalyzed by AGXT2L1 in presence of L-Cysteine or D-cysteine.

Discussion

AGXT2L1 and AGXT2L2 are not alanine-glyoxylate aminotransferases

The enzymes encoded by *AGXT2L1* and *AGXT2L2* genes show a high overall similarity (~ 55%) to human alanine-glyoxylate aminotransferase (Baker, 2004). Despite this, the present results show that AGXT2L1, and AGXT2L2 do not act physiologically as an AGT2, confirming that sequence similarity is a weak criterion for ascribing a function to a newly identified enzyme (Rost, 2002). Moreover, both enzymes did not show affinity for the AGT2 substrates, suggesting the appearance of mutations in fundamental residues involved in the substrate binding.

AGXT2L1 and AGXT2L2 ligands specificity

Recent genomic observations indicate that, in metazoa, enzyme homologs often assume non-catalytic functions – a role switch signaled by the accumulation of inactivating mutations (Pils & Schultz, 2004). In contrast to this hypothesis, however, both recombinant enzymes could still bind PLP, and react with potential substrates.

The high level of transcription of the genes in several tissues and the selectivity of recombinant enzymes toward a particular group of ligands, suggest that the AGT2-like proteins may play a biological role in the metabolism of amino acids, or amino compounds, holding a thiol group.

At least for AGXT2L1, both observations are confirmed by the finding that the enzyme catalyzes the β-elimination reactions with L-Cys and D-Cys. Clearly, AGXT2L1 reacts too slowly to be classified as a cysteine desulfhydrase (CDes). For example, if the D-cysteine desulfhydrase (D-CDes) activity for the recombinant enzyme is compared to the value obtained for D-CDes in *E.coli* (Nagasawa, 1995), AGXT2L1 is ~ $4x10^3$ times slower.

AGXT2L2 failed completely to catalyze the β-elimination reaction with L-Cys and D-Cys. Moreover, this enzymes showed a different specificity toward the same ligands of AGT2L1, hinting that substrate specificities between these two AGT2 homologs are somewhat different.

It should be considered that both these recombinant enzymes apparently bind L-Cys and D-Cys in an unproditcive way, that is, via a thiazolidine adduct. However, in principle, studying this unproductive binding could still provide substantial information about the enzymes active sites.

First of all, AGXT2L1 and AGXT2L2 seem to prefer short amino compounds holding a thiol group. Second, based on the mechanism of formation of the TZA-complex (**Scheme 2**), could be inferred that the active site environment has to permit the deprotonation of the S group in γ-position.

Evolution of AGXT2L1 and AGXT2L2

Based on the contemporary presence of AGXT2L1 and AGXT2L2 in vertebrates genomes, several speculations can be done on their function; the enzymes can catalyze the same reaction, and maybe this two form can act in different moments, or different tissues.

However, the two enzymes could have acquired different catalytic activities, after duplication in vertebrates as result of a selection pressure. In fact, during evolution, the development of new catalytic functions is believed to occur mainly by recruiting enzymes that already exist and catalyze similar reactions (O'Brien & Herschlag, 1999). In the simplest scenario, the triggering event is a duplication of the gene encoding the enzyme, which relieves the selective constraints on one gene copy, allowing for potential functional diversification. In this case, homology searches in complete genomes revealed that the AGT2-like enzymes seem derived from duplication of a single ancestor during metazoan evolution, too.

Conclusions

This work presents the first overexpression and characterization of two human recombinant alanine-glyoxylate aminotransferase type 2 homologs. Data show that both enzymes do not share substrate specificity and catalytic activity of authentic AGT2 (Noguchi, 1978; Takada, 1982; Baker, 2004; Cellini, 2007). Comparison of the recombinant proteins has also revealed catalytic and physical differences between the two proteins with regard to ligands and catalytic specificity; AGXT2L1 showed a weak activity as cysteine desulfhydrase.

Acknowledgments

This work was funded by the Telethon Foundation. I would like to thank Angelo Bolchi for the pET28-H6-Cpo vector and Lara Abdulla for molecular cloning of *AGXT2L1* and *AGXT2L2*.

References

Abe, K., S. Takahashi, et al. (2006). "Cloning and expression of the pyridoxal 5'-phosphate-dependent aspartate racemase gene from the bivalve mollusk Scapharca broughtonii and characterization of the recombinant enzyme." J Biochem 139(2): 235-44.

Akagi, R. (1982). "Purification and characterization of cysteine aminotransferase from rat liver cytosol." Acta Med Okayama 36(3): 187-97.

Baker, P. R., S. D. Cramer, et al. (2004). "Glycolate and glyoxylate metabolism in HepG2 cells." Am J Physiol Cell Physiol 287(5): C1359-65.

Brushaber, K. R., G. A. O'Toole, et al. (1998). "CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (R)-1-amino-2-propanol O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in Salmonella typhimurium LT2." J Biol Chem 273(5): 2684-91.

Canle, A., A. Lawley, E. C. McManus and R.A. More O'Ferrall (1996). "Rate and equilibrium constants for oxazolidine and thiazolidine ring-opening reactions." Pure & Appl. Chem 68(4): 813-818.

Cellini, B., M. Bertoldi, et al. (2007). "Human wild-type alanine:glyoxylate aminotransferase and its naturally occurring G82E variant: functional properties and physiological implications." Biochem J 408(1): 39-50.

Christen, P. and P. K. Mehta (2001). "From cofactor to enzymes. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Chem. Rec. 1(6): 436-447.

Coulter-Mackie, M. B. and G. Rumsby (2004). "Genetic heterogeneity in primary hyperoxaluria type 1: impact on diagnosis." Mol Genet Metab 83(1-2): 38-46.

Danpure, C. J. (2006). "Primary hyperoxaluria type 1: AGT mistargeting highlights the fundamental differences between the peroxisomal and mitochondrial protein import pathways." Biochim Biophys Acta 1763(12): 1776-84.

Donini, S., R. Percudani, et al. (2006). "A threonine synthase homolog from a mammalian genome." Biochem Biophys Res Commun 350(4): 922-8.

Fogle, E. J., W. Liu, et al. (2005). "Role of Q52 in catalysis of decarboxylation and transamination in dialkylglycine decarboxylase." Biochemistry 44(50): 16392-404.

Gophna, U., E. Bapteste, et al. (2005). "Evolutionary plasticity of methionine biosynthesis." Gene 355: 48-57.

Grishin, N. V., M. A. Phillips, et al. (1995). "Modeling of the spatial structure of eukaryotic ornithine decarboxylases." Protein Sci. 4(7): 1291-1304.

Ip, M. P., R. J. Thibert, et al. (1977). "Purification and partial characterization of cysteine-glutamate transaminase from rat liver." Can J Biochem 55(9): 958-64.

Karsten, W. E. and P. F. Cook (2002). "Detection of intermediates in reactions catalyzed by PLP-dependent enzymes: O-acetylserine sulfhydrylase and serine-glyoxalate aminotransferase." Methods Enzymol. 354: 223-237.

Keller, J. W., K. B. Baurick, et al. (1990). "Pseudomonas cepacia 2,2-dialkylglycine decarboxylase. Sequence and expression in Escherichia coli of structural and repressor genes." J Biol Chem 265(10): 5531-9.

Laber, B., K. P. Gerbling, et al. (1994). "Mechanisms of interaction of Escherichia coli threonine synthase with substrates and inhibitors." Biochemistry 33(11): 3413-3423.

Latta, K. and J. Brodehl (1990). "Primary hyperoxaluria type I." Eur J Pediatr 149(8): 518-22.

Lee, I. S., Y. Muragaki, et al. (1995). "Molecular cloning and sequencing of a cDNA encoding alanine-glyoxylate aminotransferase 2 from rat kidney." J Biochem 117(4): 856-62.

Liepman, A. H. and L. J. Olsen (2003). "Alanine aminotransferase homologs catalyze the glutamate:glyoxylate aminotransferase reaction in peroxisomes of Arabidopsis." Plant Physiol 131(1): 215-27.

Lumb, M. (2000). "Functional Synergism between the Most Common Polymorphism in Human Alanine:Glyoxylate Aminotransferase and Four of the Most Common Disease-causing Mutations*." JBC 275: 36415-36422.

Mehta, P. K. and P. Christen (2000). "The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes." Adv. Enzymol. 74: 129-184.

Meola, J. M. and H. H. Brown (1974). "Letter: Fluorescamine as a fluorogenic reagent for d-amphetamine identification by TLC." Clin Chem 20(1): 88-9.

Meyer, P., D. Liger, et al. (2005). "Crystal structure and confirmation of the alanine:glyoxylate aminotransferase activity of the YFL030w yeast protein." Biochimie 87(12): 1041-7.

Mozzarelli, A., A. Peracchi, et al. (1989). "Microspectrophotometric studies on single crystals of the tryptophan synthase a2b2 complex demonstrate formation of enzyme-substrate intermediates." J. Biol. Chem. 264(27): 15774-15780.

Nagasawa, T., T. Ishii, et al. (1985). "D-Cysteine desulfhydrase of Escherichia coli. Purification and characterization." Eur J Biochem 153(3): 541-51.

Noguchi, T., E. Okuno, et al. (1978). "Characteristics of hepatic alanine-glyoxylate aminotransferase in different mammalian species." Biochem J 169(1): 113-22.

Noguchi, T. and Y. Takada (1978). "Purification and properties of peroxisomal pyruvate (glyoxylate) aminotransferase from rat liver." Biochem J 175(2): 765-8.

Noguchi, T. (1987). "Amino acid metabolism in animal peroxisomes." Peroxisomes in Biology and Medicine. Springer-Verlag, Berlin: 234-243.

O'Brien, P. J. and D. Herschlag (1999). "Catalytic promiscuity and the evolution of new enzymatic activities." Chem. Biol. 6(4): R91-R105.

Percudani, R. and A. Peracchi (2003). "A genomic overview of pyridoxal-phosphate-dependent enzymes." EMBO Rep. 4(9): 850-854.

Pils, B. and J. Schultz (2004). "Inactive enzyme-homologues find new function in regulatory processes." J. Mol. Biol. 340(3): 399-404.

Ponticelli, F., R. Pagani, et al. (1983). "Synthesis and NMR properties of 2-(4-pyridyl) thiazolidine-4-carboxylic acids." Ital J Biochem 32(2): 92-101.

Rost, B. (2002). "Enzyme function less conserved than anticipated." J. Mol. Biol. 318(2): 595-608.

Rowsell, E. V., K. Snell, et al. (1969). "Liver-L-alanine-glyoxylate and L-serine-pyruvate aminotransferase activities: an apparent association with gluconeogenesis." Biochem J 115(5): 1071-3.

Rowsell, E. V., A. H. al-Tai, et al. (1972). "Liver L-serine-pyruvate aminotransferase activity in different animal species." Biochem J 127(2): 27P.

Shen, B. W., M. Hennig, et al. (1998). "Crystal structure of human recombinant ornithine aminotransferase." J Mol Biol 277(1): 81-102.

Sterk, M. and H. Gehring (1991). "Spectroscopic characterization of true enzyme-substrate intermediates of aspartate aminotransferase trapped at subzero temperatures." Eur. J. Biochem. 201(3): 703-7.

Storici, P., G. Capitani, et al. (1999). "Crystal structure of GABA-aminotransferase, a target for antiepileptic drug therapy." Biochemistry 38(27): 8628-34.

Storici, P., D. De Biase, et al. (2004). "Structures of gamma-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5'-phosphate, and [2Fe-2S] cluster-containing enzyme, complexed with gamma-ethynyl-GABA and with the antiepilepsy drug vigabatrin." J Biol Chem 279(1): 363-73.

Takada, Y. and T. Noguchi (1982). "Subcellular distribution, and physical and immunological properties of hepatic alanine: glyoxylate aminotransferase isoenzymes in different mammalian species." Comp Biochem Physiol B 72(4): 597-604.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res. 25(24): 4876-4882.

Wolosker, H., K. N. Sheth, et al. (1999). "Purification of serine racemase: biosynthesis of the neuromodulator D-serine." Proc Natl Acad Sci U S A 96(2): 721-5.

Zhang, X., S. M. Roe, et al. (2003). "Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1." J Mol Biol 331(3): 643-52.

Appendix

The advantage of being locked

Introduction

In the past decade, several RNA-cleaving deoxyribozymes (DNAzymes)[4] have been isolated in the laboratory through the application of a combinatorial technique known as "*in vitro* selection" (Silverman, 2005). The two most prominent examples are the 8-17 and 10-23 deoxyribozymes, originally identified in the Joyce lab (Santoro and Joyce, 1997). Both types of deoxyribozymes comprise a central "core" and two substrate-binding arms that can be varied in length and sequence.

The chemical stability of these catalysts, their ease of synthesis and versatility in the sequence-specific cleavage of RNA have prompted a wealth of studies on the potential use of DNAzymes in therapy, to inhibit the expression of disease-causing genes by selectively binding and cleaving the corresponding mRNAs (Cairns, 2002; Dass, 2004; Dass, 2006).

However, several factors limit the practical use of DNAzymes as anti-gene agents. For example, it may be difficult to achieve an intercellular concentration of DNAzyme high enough to ensure stable binding to the target mRNA and hence efficient knockdown. In part, this problem can be countered by lengthening the substrate-recognition arms, thus increasing the affinity of the DNAzyme for its target, but this entails the synthesis of larger DNA molecules, less manageable and more prone to the adoption of alternative structures and to off-target binding.

The binding problem is made even worse by the potentially limited accessibility of the intended target sequence within the mRNA structure. Indeed, there are suggestions that up to 90% of potential cleavage sites in long mRNAs are inaccessible to classical DNAzymes (Cairns, 1999; Kurreck, 2002) presumably due to a complex secondary and tertiary folding of these RNAs.

An important step forward in solving these problems of efficient binding at low DNAzyme concentrations and of target accessibility has been the introduction of LNA (locked nucleic acid; **Figure 1A**) LNA residues are capable of normal Watson-Crick basepairing, and the locked sugar enhances the preorganization of the phosphate backbone, stabilizing the helical (A-type) structure of a duplex (Vester, 2004; McTigue, 2004; Kaur, 2006).

[4] The abbreviations used are: DNAzyme, a deoxyribozyme construct composed entirely of DNA; LNA, locked nucleic acid; LNAzyme, a deoxyribozyme construct containing some LNA nucleotides in the substrate-binding arms; HPV16, human papillomavirus type 16; PIPES, 1,4-piperazinediethansulfonic acid; nt, nucleotide.

Several studies have shown that the inclusion of a few LNA monomers into the arms of the 10-23 deoxyribozyme significantly improves the performances of this catalyst against small RNA substrates, and, more pronouncedly, against long RNA transcripts, promoting cleavage of highly structured RNAs (Vester, 2002; Schubert, 2003; Fahmy, 2004; Schubert, 2004; Fluiter, 2005; Vester, 2006; Jakobsen, 2007).

Although LNA monomers are expected to increase the affinity of a deoxyribozyme for its target, the kinetics and thermodynamics underlying the observed catalytic enhancements have not been analyzed in detail. With respect to a minimal kinetic scheme for a deoxyribozyme reaction (**Figure 1B**), it remains to be established whether the increased affinity alone can explain the improved kinetics, or if LNAs also accelerates the intramolecular cleavage step (described by the kinetic constant k_2) (Vester, 2002; Schubert, 2003; Fahmy, 2004; Schubert, 2004; Fluiter, 2005; Vester, 2006).

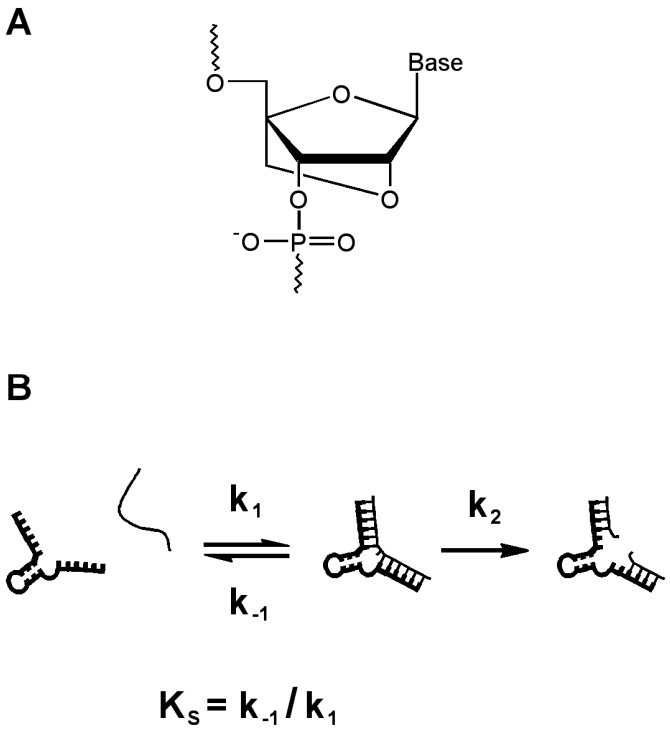


Figure 1. (A) Structure of an LNA nucleotide. (B) A minimal kinetic scheme for the RNA cleavage reaction carried out by a DNAzyme (or LNAzyme), under single-turnover conditions. The catalyst (drawn with a dark line) must bind its substrate (thin line) to form a productive biomolecular complex. Within this complex, the target RNA is cleaved to yield two products, one of which containing a cyclic 2',3'-phosphate (Santoro, 1997).

Furthermore, the LNA monomers could in principle improve the RNA binding affinity of deoxyribozymes *via* an increased association rate (k_1) or a reduced

dissociation rate (k_{-1}), or both; data from model systems are somewhat contradictory on this point (Christensen, 2001; Ormond, 2006). Finally, it is not completely obvious why the improved performances of LNA-containing deoxyribozymes were particularly evident against long RNA transcripts.

In relation to these issues, we provide here a detailed kinetic comparison between a well-characterized 8-17 construct and a corresponding LNA-armed version. Furthermore, we contrast the performances of four DNAzyme/LNAzyme pairs targeted against different regions of a viral RNA transcript (the E6 mRNA from HPV16). Our findings indicate that (i) LNA monomers located in the deoxyribozyme arms have only minimal effects on the cleavage step (k_2); (ii) LNA monomers greatly reduce the rate of substrate dissociation (k_{-1}), stabilizing the catalyst-substrate complex; (iii) finally, the effect on the association step (k_1) depends on the type of substrate: with short, unstructured substrates k_1 is nearly equal for DNAzymes and LNAzymes, whereas with long, structured RNAs, association is substantially faster for LNAzymes as compared to DNAzymes.

Materials and Methods

Oligonucleotides

DNA oligonucleotides were from MWG Biotech (Ebersberg, Germany) or from DNAtechnology (Aarhus, Denmark). LNAzymes were synthesized using published procedures (Wengel, 1999). The minimal (17-mer) RNA substrate for Dz122 and Lz122 was from Dharmacon Research (Lafayette, CO) and when necessary it was ^{32}P-5'-end labeled with T4 polynucleotide kinase. The other minimal substrates were synthesized with T7 RNA polymerase (Promega), employing the following templates:

S32 Template
$^{5'}$GTGGTAACTTTCTGGGTCCTATAGTGAGTCGTATTAG$^{3'}$

S51 Template
$^{5'}$TGCAGCTCTGTGCATAACCTATAGTGAGTCGTATTAG$^{3'}$

S122 Template:
$^{5'}$CATATACCTCACGTCGCCCTATAGTGAGTCGTATTAG$^{3'}$

S163 Template:
$^{5'}$CCCATCTCTATATACTACCTATAGTGAGTCGTATTAG$^{3'}$

The templates were hybridized to the complementary deoxyoligonucleotide to form a double-stranded T7 promoter.

T7 promoter:

$^{5'}$ TAATACGACTCACTATAGG$^{3'}$

Transcription was conducted using the buffer conditions recommended by the supplier, and 1 mM nucleoside triphosphates. After transcription, the reaction mixtures were extracted and precipitated and the full-length RNAs were isolated on a 7M urea/13% polyacrylamide gel. RNAs were eluted from gel bands, extracted with phenol/chloroform, ethanol-precipitated and redissolved in H_2O.

RNA transcripts were generally 5´-radiolabeled by dephosphorylating using shrimp phosphatase (USB) and subsequently incubating with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP. Concentrations of radioactive

oligonucleotides were determined from specific activities; concentrations of nonradioactive oligonucleotides were determined using extinction coefficients at 260 nm estimated by the nearest neighbor method (Cantor, 1970).

In vitro transcription of E6 and E6E7 mRNA

The E6 and E6E7 mRNAs from HPV16 were transcribed *in vitro* from the corresponding genes. Two plasmids containing the E6 coding sequence (BSK-16-E6) and the E6E7 coding sequence (BKS-16-E6E7) were kindly provided by Massimo Tommasino (International Agency for Research on Cancer, Lyon). These were Bluescript plasmids (the complete sequence and list of restriction sites are available at www.stratagene.com - Genbank® #X52327[KS(+)] and #52328 [SK(+)]).

The plasmids were first cut with XhoI (a restriction enzyme whose cleavage site is located, in both constructs, shortly downstream of the stop codon – see **Figure 2**) and then transcribed using either T3 RNA polymerase (for the E6 RNA) or T7 RNA polymerase (for the E6E7 RNA). The transcription mixture (50 µl) for the non-radioactive substrates contained: 10 µl 5 x transcription buffer, 7 mM nucleoside triphosphate, 26 mM $MgCl_2$, 30 mM DTT, 750 ng of template and 17 U T3 RNA polymerase (Amersham), or 20U T7 RNA polymerase (Stratagene). The transcription mixture (20 µl) for the radioactive substrate contained: 4 µl 5 x transcription buffer, 1mM nucleoside triphosphate, 10 mM $MgCl_2$, 10 mM DTT, 29 U RNAsin, 2 µg of template and polymerases as above.

After transcription, the RNAs were extracted with phenol-chloroform, precipitated and purified on a 7M urea/6% polyacrylamide gel. Full-length transcripts were eluted from gel bands using 2 M NH_4Ac, pH 5.3 (elution overnight with 150 µl and wash with 100 µl), extracted with phenol/chloroform, ethanol-precipitated and redissolved in H_2O. Concentrations of radioactive and nonradioactive oligonucleotides were determined using extinction coefficients at 260 nm estimated by the nearest neighbor method (Cantor, 1970).

A

E6 Transcript

GGGAACAAAAGCUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAACUAGUGGAUCCCCCGGG**AUG**U
UUCAGGACCCACAGGAGCGACCCAGAAAGUUACCACAGUUAUGCACAGAGCUGCAAACAACUAUAC
AUGAUAUAAUAUUAGAAUGUGUGUACUGCAAGCAACAGUUACUGCGACGUGAGGUAUAUGACUUUG
CUUUUCGGGAUUUAUGCAUAGUAUAUAGAGAUGGGAAUCCAUAUGCUGUAUGUGAUAAAUGUUUAA
AGUUUUAUUCUAAAAUUAGUGAGUAUAGACAUUAUUGUUAUAGUUUGUAUGGAACAACAUUAGAAC
AGCAAUACAACAAACCGUUGUGUGAUUUGUUAAUUAGGUGUAUUAACUGUCAAAAGCCACUGUGUC
CUGAAGAAAAGCAAAGACAUCUGGACAAAAAGCAAAGAUUCCAUAAUAUAAGGGGUCGGUGGACCG
GUCGAUGUAUGUCUUGUUGCAGAUCAUCAAGAACACGUAGAGAAACCCAGCUG**UAA**AAGCUUAUCG
AUACCGUCGACCUCGA

B

E6E7 Transcript

GGGCGAAUUGGAGCUCCACCGCGGUGGCGGCCGCUCUAGAACUAGUGGAUCCCCCGGGCUGCAGGA
AUUC**AUG**UUUCAGGACCCACAGGAGCGACCCAGAAAGUUACCACAGUUAUGCACAGAGCUGCAAAC
AACUAUACAUGAUAUAAUAUUAGAAUGUGUGUACUGCAAGCAACAGUUACUGCGACGUGAGGUAUA
UGACUUUGCUUUUCGGGAUUUAUGCAUAGUAUAUAGAGAUGGGAAUCCAUAUGCUGUAUGUGAUAA
AUGUUUAAAGUUUUAUUCUAAAAUUAGUGAGUAUAGACAUUAUUGUUAUAGUGUGUAUGGAACAAC
AUUAGAACAGCAAUACAACAAACCGUUGUGUGAUUUGUUAAUUAGGUGUAUUAACUGUCAAAAGCC
ACUGUGUCCUGAAGAAAAGCAAAGACAUCUGGACAAAAAGCAAAGAUUCCAUAAUAUAAGGGGUCG
GUGGACCGGUCGAUGUAUGUCUUGUUGCAGAUCAUCAAGAACACGUAGAGAAACCCAGCUGUAAUC
AUGCAUGGAGAUACACCUACAUUGCAUGAAUAUAUGUUAGAUUUGCAACCAGAGACAACUGAUCUC
UACUGUUAUGAGCAAUUAAAUGACAGCUCAGAGGAGGAGGAUGAAAUAGAUGGUCCAGCUGGACAA
GCAGAACCGGACGGAGCCCAUUACAAUAUUGUAACCUUUUGUUGCAAGUGUGACUCUACGCUUCGG
UUGUGCGUACAAAGCACACACGUAGACAUUCGUACUUUGGAAGACCUGUUAAUGGGCACACUAGGA
AUUGUGUGCCCCAUCUGUUCUCAGAAACCA**UAA**CUCGA

Figure 2 – The long RNA transcripts employed in this work. Boxed sequences correspond to the targets of the four 8-17 constructs described in the text. (A) The E6 RNA was transcribed with T3 RNA polymerase, using the linearized BSK-16-E6 plasmid as a template. The plasmid was linearized with XhoI. The sequence corresponding to the XhoI target site is underlined, whereas the start and stop triplets of the E6 coding sequence are shown in bold. (B) The E6E7 transcript was generated using T7 RNA polymerase and the XhoI-linearized BKS-16-E6E7 plasmid as a template. The underlined sequence signals the XhoI target site; bold letters indicate the start codon of the E6 coding sequence and the stop codon of the E7 coding sequence.

Measurement of individual rate constants for Dz122 and Lz122.

Reactions of Dz122 and Lz122 against a minimal substrate (17-nt long) were conducted in 50 mM PIPES-NaOH, pH 7.4, 3 mM Mg^{2+}, to allow a direct comparison with an earlier study (Bonaccio, 2004) where the Dz122 construct was characterized in detail.

Cleavage reactions were single-turnover, with the catalyst in large excess (≥10-fold) with respect to the labeled substrate, so that the product dissociation step did not affect the observed kinetics (Bonaccio, 2004). Substrate and catalyst were separately heated at 95 °C for 2 min to disrupt potential aggregates, spun briefly in a microfuge and equilibrated for 10-30 min at the reaction temperature. After supplementing the catalyst tube with $MgCl_2$, reactions were initiated by adding the substrate. Time-points were collected at appropriate intervals and further reaction was quenched by adding formamide and excess EDTA. Radiolabeled substrates and products were separated on 7 M urea/20% polyacrylamide gels and quantitated by phosphorimaging. Reaction time-courses were fit to the appropriate kinetic equation using Sigma Plot (SPSS Inc.).

k_2 (describing cleavage within the enzyme-substrate complex) was measured employing saturating concentrations of the catalyst (≥ 1 μM), to ensure that all the substrate would be bound to the DNAzyme or LNAzyme (Bonaccio, 2004). As a control that saturation of the substrate was achieved, the measured rate constants remained identical, within error, when the concentration of catalyst was raised from 1 μM to 5 μM or 10 μM.

The rate constants for substrate dissociation (k_{-1}) and association (k_1) were measured through a pulse-chase strategy (Bonaccio, 2004). To determine k_{-1}, a saturating concentration of DNAzyme or LNAzyme (100 nM) was first allowed to bind a trace amount of radiolabeled substrate (~0.1 nM) for 4 min, in the presence of 3 mM $MgCl_2$ (over this time, <1% of the substrate was cleaved). Then a large excess of unlabeled substrate (1 μM final) was added to initiate the ‘chase‘ period, during which dissociation of labeled substrate from the catalyst was essentially irreversible. An otherwise identical reaction, but without the chase, was carried out in parallel. Partitioning of the labeled substrate between cleavage and release depended on the relative magnitude of k_{-1} and k_2; in particular, the fractional extent of cleavage in the presence of the chase (relative to the extent in the absence of chase) reflected the

ratio $k_2/(k_{-1}+k_2)$ (Fersht, 1985). Thus, measuring k_2 and the final extent of cleavage allowed determination of k_{-1}.

A pulse-chase method was also used to determine k_1. In these experiments, an excess DNAzyme or LNAzyme (1-15 nM) was first allowed to bind a trace amount of radiolabeled substrate for a period, t_1, variable from 0.5 to 7 min (over this time, <1% of the substrate was cleaved). Then the sample was transferred into a solution containing a large excess of unlabeled substrate (to prevent any further binding of radioactive substrate) as well as 15 mM Mn^{2+}, pH 7.8.

Due to the high concentration of Mn^{2+} and the increased pH, all the bound labeled substrate was cleaved in less than 3 min (Bonaccio, 2004), so that the fraction of cleaved substrate reflected the amount of catalyst-substrate complex formed at t_1. The dependence of such fraction on t_1 was fit to a single exponential function, yielding a pseudo-first order rate constant for substrate binding, k_{app}. k_1 was obtained from a linear regression of the k_{app} values vs. the catalyst concentration.

Cleavage of long RNA transcripts

Cleavage of the E6 and E6E7 mRNAs were measured under single-turnover conditions at 37°C. The reaction buffer contained 50 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.008% SDS.

Two distinct kinetic protocols were used. In protocol A, the DNAzyme (or LNAzyme) and the mRNA substrate were heated separately in reaction buffer, in the presence of 10 mM $MgCl_2$. After this heating step (2 min at 80°C), the samples were spun briefly in a microfuge and equilibrated for 10 min at the reaction temperature. The cleavage reactions were initiated by mixing the catalyst and substrate solutions. The reaction mixture contained 10 nM RNA (final concentration) and 100 nM catalyst (tenfold molar excess). Aliquots were collected at appropriate times and quenched by adding formamide and excess EDTA.

In protocol B, cleavage rates were determined after pre-annealing the catalyst and the mRNA. Annealing was achieved by heating together the DNAzyme (or LNAzyme) and RNA in reaction buffer at 80°C for two minutes, followed by cooling in two steps (50°C for 5 min; 37°C for 10 min). The cleavage reactions were then

started by adding $MgCl_2$ to a final concentration of 10 mM. The concentrations of enzyme and RNA were as above.

Substrates and products of the cleavage reactions were separated on 6% polyacrylamide gels and quantitated by a phosphorimager. The fraction of product at time t, F_t, was calculated by dividing the amount of product by the amount of substrate plus product and the data were fit to a single-exponential function:

$$F_t = F_\infty\left(1 - e^{-k_{obs}t}\right) \qquad \text{(Eq. 1)}$$

Where F_∞ is the fraction of product at the endpoint of the reaction and k_{obs} is the rate constant of cleavage (Fersht, 1985). Sometimes the reaction was clearly biphasic, and the data were hence fit to a double exponential function:

$$F_t = F_1\left(1 - e^{-k_{obs1}t}\right) + F_2\left(1 - e^{-k_{obs2}t}\right) \qquad \text{(Eq. 2)}$$

Where F_1 and F_2 represent the endpoints of the first and second phase of the time-course, respectively.

Results

Constructs used

In this study, we compared the activities of four conventional 8-17 DNAzymes with their partially LNA-armed counterparts ('LNAzymes'; **Figure 3**). All four 8-17 constructs were directed against the E6 mRNA from Human Papillomavirus type 16 (HPV16). This mRNA is a well-established model system to explore the application of antisense and catalytic antisense oligonucleotides and much experimental information is available about the accessibility of different parts of the RNA sequence (Cairns, 1999; Venturini, 1999; Pan, 2003; Cairns and Sun, 2004).

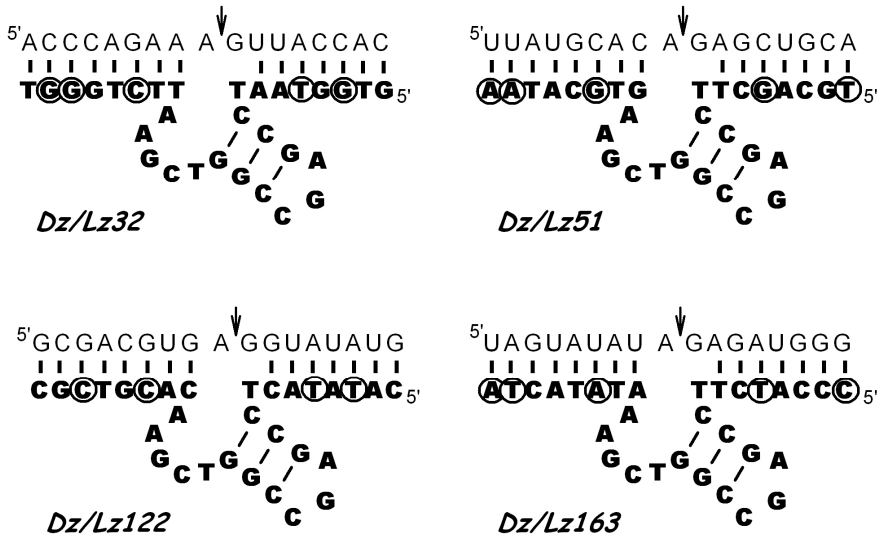


Figure 3. Primary and secondary structures of the 8-17 DNAzymes and LNAzymes used in this study. The four constructs were designed to recognize 17-nt-long RNA targets on the E6 RNA. The number in each construct's name refers to the position of cleavage within the E6 coding sequence. Against each target, we tested one 8-17 DNAzyme (composed only of DNA monomers) and one LNAzyme (in which four or five nucleotides in the arms were replaced by LNA monomers). DNA is shown in bold letters, and the positions at which LNA monomers were introduced are circled. Note that Cytidine LNA residues are methylated at the 5-position. The RNA target sequences are shown as normal letters while the arrows indicate the cleavage sites.

The four 8-17 constructs were designed to cleave specific AG sites, located within the first 200 nucleotides of the E6 coding sequence. Dz32 was designed to hybridize to a sequence that overlaps the best target sequence identified by Pan et al., who tested the accessibility of the E6 mRNA by using a library of hammerhead ribozymes (Pan, 2003). The Dz122 target sequence includes two of the four best cleavage sites identified by Cairns and co-workers, using a library of 10-23

deoxyribozymes (Cairns, 1999). The target of Dz163 encompassed one accessible cleavage site identified by Pan et al. and two accessible sites described by Cairns et al. (Cairns, 1999; Pan, 2003). Finally, Dz51 was used as a control; it targets a sequence that previous studies found only modestly accessible (Cairns, 1999; Pan, 2003), but RNA folding software such as mFold (Zuker, 2003) predicted that the cleavage site would lie in a single-stranded region.

Kinetic and thermodynamic comparison of Dz122 and Lz122

The construct Dz122 had been kinetically well characterized in previous studies, where it was named (8-17)*cb* (Bonaccio, 2004; Peracchi, 2005; Ferrari, 2002). Therefore we employed this construct for a detailed kinetic and thermodynamic comparison with the corresponding LNAzyme (termed Lz122), using a short 17mer RNA substrate.

The rate constant for cleavage of the 17-mer RNA within the enzyme-substrate complex (k_2) was measured in single-turnover reactions (ensuring that the product dissociation step did not affect the observed kinetics) and at saturating deoxyribozyme concentration (≥ 1 μM; Bonaccio, 2004). Dz122 and Lz122 showed very similar k_2 values, both at 25°C and 37°C (**Table 1**), indicating that the introduction of the four LNA monomers in the 8-17 arms has only marginal effects on the intramolecular strand-scission step.

Table 1
Comparison of the catalytic parameters for the reactions catalyzed by Dz122 and Lz122.

Construct	Temperature	k_2 (min^{-1})	k_1 ($M^{-1}min^{-1}$)	k_{-1} (min^{-1})	K_S (nM)
Dz122	25°C	10^{-2}	1.7×10^{7}	0.02	1.2
Lz122	25°C	6×10^{-3}	3.3×10^{7}	0.0016	0.05
Dz122	37°C	2.8×10^{-2}	-	-	-
Lz122	37°C	2.2×10^{-2}	-	-	-

Table 1. Conditions: PIPES-NaOH buffer, pH 7.4, 3 mM Mg^{2+}. Data are the average of at least two independent determinations, which differed by less than 25% from each other. The k_1 and k_{-1} values for the Dz122 construct had already been reported in Bonaccio, 2004.

The measured association rate constants (k_1) also showed differences of less than twofold between Dz122 and Lz122 (**Table 1**), in agreement with previous indications that LNAs have only minimal effects on the hybridization rate between short oligonucleotides (Christensen, 2001). Much larger differences were observed when measuring k_{-1}, the rate constant for dissociation of the enzyme-substrate complex. The values of k_{-1} for Dz122 and Lz122 differed by more than tenfold, reflecting the different stabilities of binding for the two catalysts. From the kinetic constants k_1 and k_{-1} we could also calculate the thermodynamic constant for substrate dissociation ($K_S = k_{-1}/k_1$), which was ~25-fold lower for the LNAzyme (**Table 1**).

These data were buttressed by UV melting studies on the complexes formed by the enzymes with a non-cleavable substrate analog, bearing a single deoxyribonucleotide at the cleavage site. The experiments showed that the melting temperature (T_m) for the complex between Lz122 and the substrate analog was ≈20°C higher than for the Dz122 complex (58°C vs. 38°C, using 2 μM catalyst and 2 μM substrate analog; **Figure 4**).

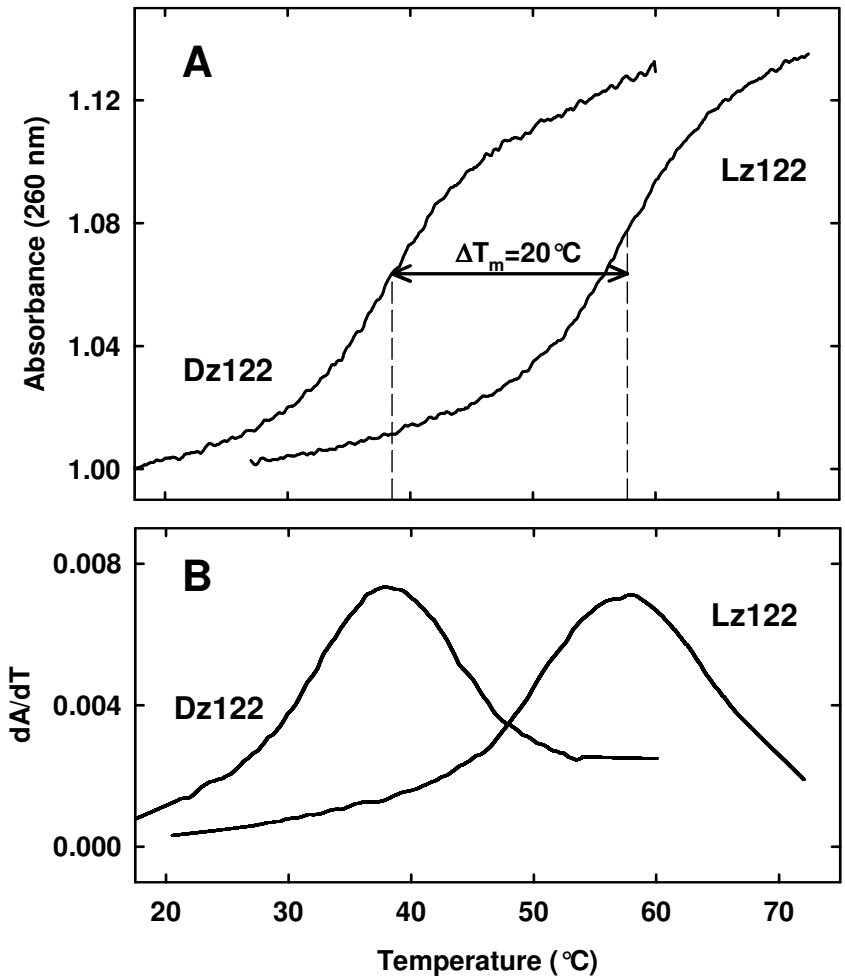


Figure 4. Thermal denaturation of the complexes formed between Dz122 or Lz122 and a substrate analog containing a single deoxyribonucleotide at the cleavage site. Denaturation was monitored in PIPES-NaOH buffer, pH 7.4 (at 37°C), 3 mM $MgCl_2$. In each experiment, catalyst and substrate analog were both present at 2 μM concentration. Measurements were performed on a Perkin-Elmer Lambda Bio 20 spectrophotometer, equipped with a PTP-6 Peltier temperature programmer. Enzyme and substrate analog were heated together at 95 °C for 2 min, then cooled at room temperature, supplemented with 3 mM $MgCl_2$ and finally placed into the spectrophotometer and equilibrated at 10 °C. Melting curves (absorbance at 260 nm vs. temperature) were recorded by heating the samples (1°C/minute) and following the absorbance variation at 260 nm. The exterior of the cuvette was kept under a stream of dry N_2 gas to prevent water condensation at low temperatures. (A) Melting curves. The ordinate represents the relative change of absorbance as a function of temperature. (B) First derivative of UV absorption with respect to temperature. The two graphs for Dz122 and Lz122 show maxima near (respectively) 38°C and 58°C.

These findings are in agreement with published data, which report that the T_m of short DNA-RNA duplexes is increased by 4 to 7°C per each pyrimidine LNA monomer introduced in the DNA sequence (Singh, 1998; Singh, 1998, Bondensgaard, 2000). The results also indicate that, length and sequence of the binding arms being equal, an LNAzyme can be employed at a range of temperatures much wider than its all-DNA counterpart (**Figure 5**).

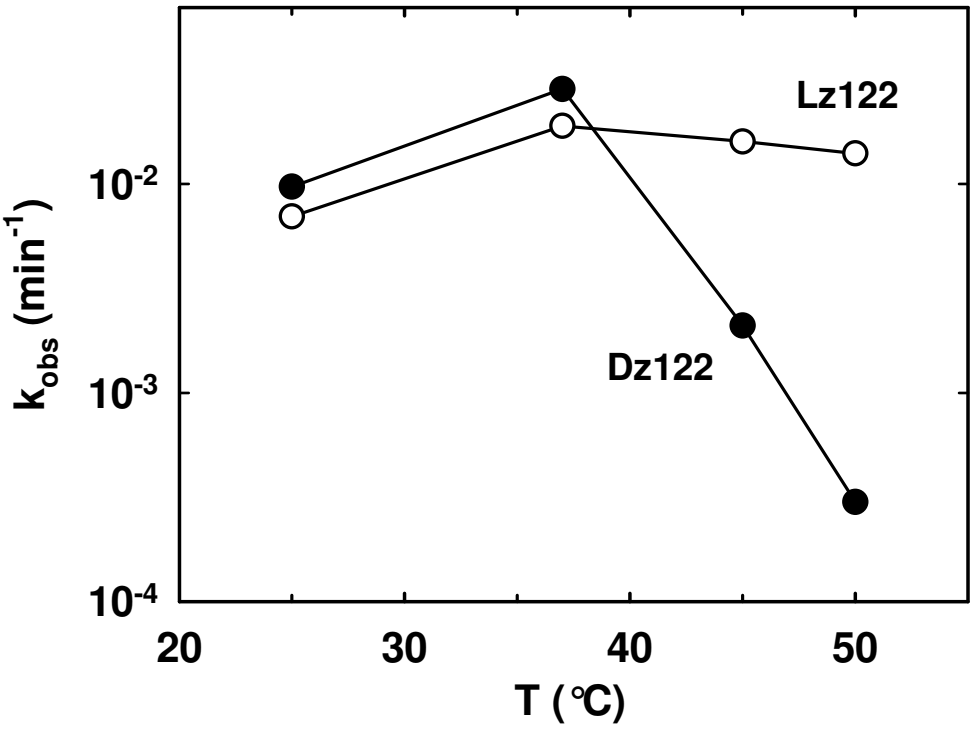


Figure 5 – Temperature dependence of the cleavage rates observed with Dz122 and Lz122, under single turnover conditions, at pH 7.4, 3 mM Mg^{2+}. The concentration of catalyst was 1 μM. The decrease in activity of Dz122 at T>37°C reflects the dissociation of the DNAzyme from its substrate (see preceding figure) and the ensuing loss of catalytic efficiency.

Activities of Dz122 and Lz122 as a function of catalyst concentration and Mg^{2+} concentration

The higher thermodynamic stability of the LNAzyme-substrate complex, coupled with the minimal effects of LNA on the cleavage step, imply that LNAzymes can outperform DNAzymes under a number of conditions. For example, when measuring substrate cleavage as a function of catalyst concentration, at 37°C, Lz122 was substantially more efficient than Dz122 at concentrations <50 nM (**Figure 6A**). Presumably, the lower K_S of Lz122 allowed this catalyst to saturate its substrate more easily.

We also compared the Mg^{2+}-dependences of activity for Dz122 and Lz122. It is known that the intramolecular reaction rate (k_2) of 8-17 is stimulated by Mg^{2+}, and that such activation follows a hyperbolic titration curve (Bonaccio, 2004); also, the thermodynamic substrate dissociation constant, K_S, is expected to decrease when the concentration of Mg^{2+} increases (Nakano, 1999). When we tested Dz122 and Lz122 (at a fixed 100 nM concentration) vs. [Mg^{2+}], the LNAzyme activity showed a hyperbolic dependence, while the curve for Dz122 was sigmoidal, indicating an apparently cooperative activation (**Figure 6B**).

As a result, the LNAzyme was significantly superior to the DNAzyme at [Mg^{2+}]≤1.5 mM (**Figure 6B**). We explain these findings as follows. For Lz122, the 100 nM catalyst concentration is sufficient to saturate the substrate over the whole range of [Mg^{2+}] explored, so that only the effect on k_2 is apparent. Dz122, however, is unable to saturate its substrate at low magnesium concentrations, hence performing with suboptimal efficiency. The apparent cooperativity arises because Mg^{2+} ions, at low concentrations, stimulate both formation of the Dz122-substrate complex (effect on K_S) and catalysis within the complex (effect on k_2).

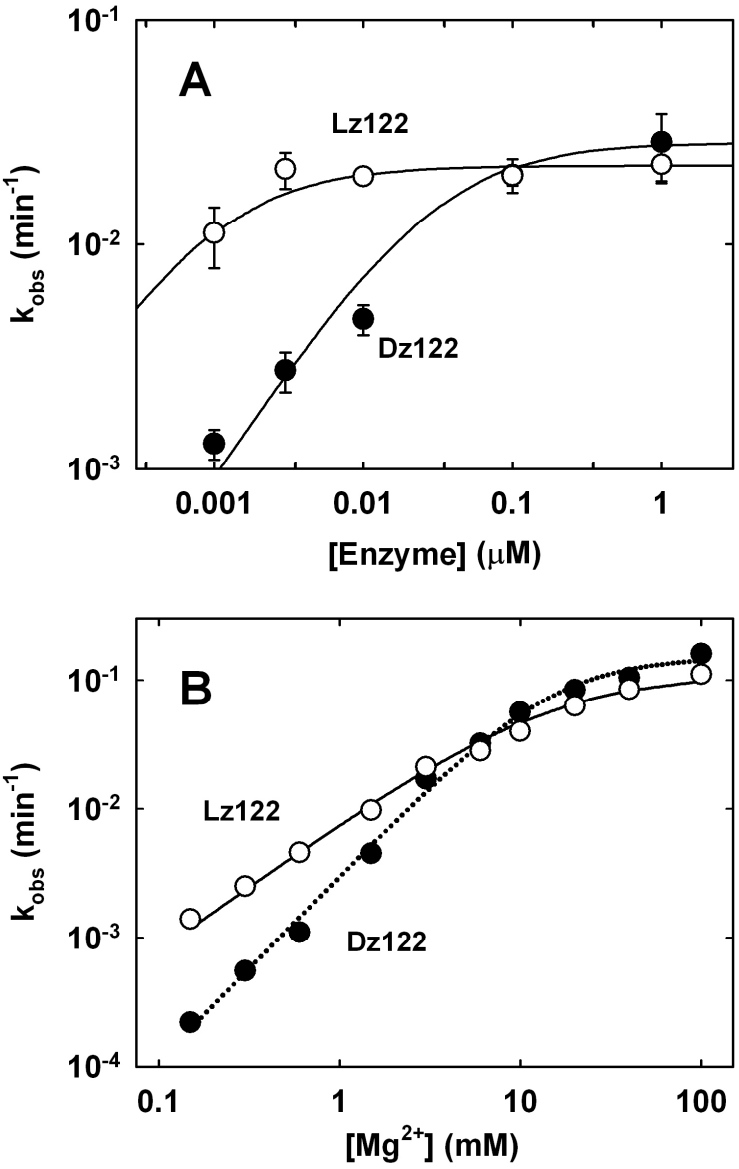


Figure 6. (A) Concentration dependence of the cleavage rates observed with Dz122 and Lz122 at 37°C, pH 7.4, 3 mM Mg^{2+}. Error bars represent the range of variation in duplicate experiments. The data were fit to the following equation:

$$k_{obs} = \frac{k_2 \times [E]}{K_s + [E]} \qquad \text{(Eq. 3)}$$

where [E] represents the catalyst concentration. The solid lines through the data represent binding curves with K_S=30 nM (Dz122) and 1 nM (Lz122). The estimated 30-fold greater affinity of Lz122 for its substrate compares well with the 24-fold difference obtained at 25°C (**Table 1**). (B) Mg^{2+}-dependence of the cleavage rates. Conditions as in panel A, except that the concentration of the catalyst was constant (100 nM) and the concentration of Mg^{2+} varied from 0.15 to 100 mM. The solid line trough the Lz122 data is the best fit to a binding hyperbola with K_d=14 mM. The dotted line through the Dz122 data is the best fit to the Hill equation (Fersht, 1985). The Hill coefficient, n, gauging the apparent cooperativity of Mg^{2+} activation, was 1.4.

LNAzymes, but not DNAzymes, cleave efficiently the folded E6 and E6E7 transcripts

We next compared the relative cleavage efficiencies of 8-17 DNAzymes and LNAzymes against targets that are embedded in large, structured RNA molecules. To this end, we assayed the ability of the four DNAzymes and of the respective LNAzymes to cleave the full-length E6 transcript (E6 mRNA, 544 nt long). Table 2 summarizes the results obtained with this long transcript, together with controls where minimal (19-mer) substrates were used. Note that, during HPV16 infection, the E6 sequence is part of a polycistronic mRNA that also encodes the E7 oncoprotein; accordingly, as a further control, we tested the DNAzymes and LNAzymes also against an E6E7 polycistronic transcript (E6E7 mRNA 830 nt long; Supplemental Figure 1).

Reaction rates were initially measured using 'protocol A' (see Experimental Procedures). In this protocol, mRNA transcripts are allowed to fold into stable structures (presumably similar to the structures adopted in the cytoplasm) prior to being exposed to the catalysts. Under these conditions, the LNAzymes (at a 100 nM concentration) could efficiently cleave the long RNAs (**Figure 7A**), yielding kinetic time-courses that were well described by simple exponential functions (**Figure 7B**). The observed rate constants closely resembled those measured for the cleavage of minimal substrates (**Table 2**, left). Moreover, the rates of cleavage were virtually unchanged when the four LNAzymes were tested against the E6 mRNA or the E6E7 mRNA, in spite of the additional ~300 nt in the E6E7 transcript (**Table 3**).

While three out of four LNAzymes showed similar activities against the long RNAs, Lz51 cleaved much more slowly and to a lower extent (**Figure 7B**). As stated above, previous studies implied that the target sequence of Lz51 could be intrinsically less accessible than those of the other three constructs, but this would not explain why Lz51 behaved poorly even against a short 19-mer substrate (**Table 2**). A more likely explanation lies in the propensity of free Lz51 to adopt alternative secondary structures, that are expected to reduce binding of the substrate and hence catalysis (data not shown). In agreement with this hypothesis, control experiments suggested that the LNAzyme, at the 100 nM concentration used, was far from saturating its E6 substrate (data not shown). Despite these problems, it is remarkable that Lz51 was able to access and cleave its target within the long RNAs, whereas the corresponding all-DNA construct, Dz51, failed to cleave at concentrations as high as 10 μM (**Table 2** and data not show).

Indeed, in strong contrast to the LNAzymes, all four of the DNAzymes showed a nearly complete inability to cleave the long substrates under the 'protocol A' conditions (**Table 2**, left). This suggests that the local RNA structure prevents the access of unmodified 8-17 DNAzymes to fully folded targets, exacerbating the differences in activity between DNAzymes and LNAzymes.

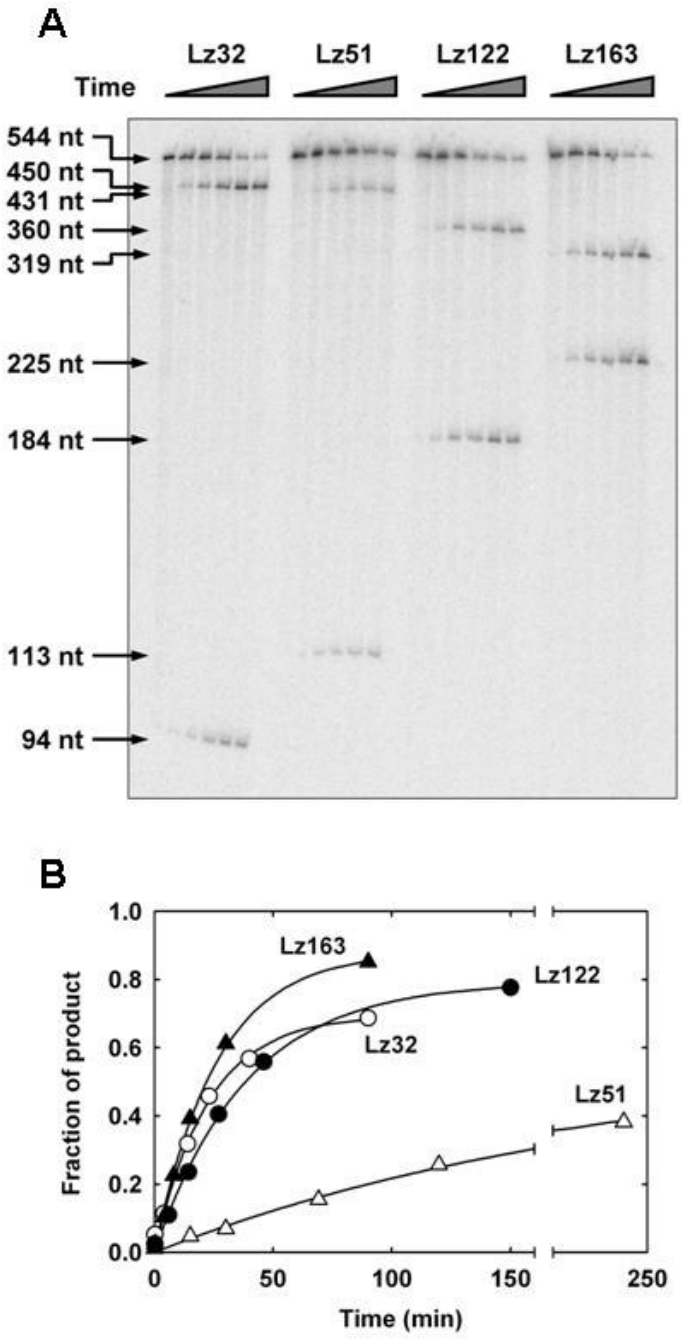


Figure 7. (A) Single turnover reaction of the LNAzymes under 'protocol A' conditions. The internally-labeled E6 transcript (544 nt long; 10 nM final concentration) was reacted with each of the four LNAzymes (100 nM) at 37°C, pH 7.5, 10 mM Mg^{2+}. Aliquots were collected at appropriate times and the substrate and products were separated on a 5% polyacrylamide denaturing gel. (B) Time-courses of the cleavage reactions shown in panel A. The solid lines represent least-squares fits of the experimental data to Equation 1.

Table 2
Performance of DNAzymes and LNAzymes against the E6 mRNA transcript

	Protocol A (cleavage of pre-folded RNA)				Protocol B (cleavage after pre-annealing catalyst and substrate)			
	19-mer substrate		E6 mRNA		19-mer substrate		E6 mRNA	
	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞
Lz32	4.6×10^{-2}	0.79	4.4×10^{-2}	0.64	6.6×10^{-2}	0.73	4.5×10^{-2}	0.67
Lz51 [a]	3.6×10^{-3}	0.8	5.5×10^{-3}	0.44	2.6×10^{-3}	0.50	1^{st}: 5.2×10^{-2} 2^{nd}: 2.0×10^{-3}	F_1=0.12 F_2=0.64
Lz122 [b]	2.0×10^{-2}	0.68	2.5×10^{-2}	0.80	3.7×10^{-2}	0.74	4.1×10^{-2}	0.75
Lz163 [c]	2.9×10^{-2}	0.85	3.2×10^{-2}	0.85	0.14	0.76	5.2×10^{-2}	0.86
Dz32 [d]	1.6×10^{-2}	0.90	$\leq10^{-4}$	-	2.1×10^{-2}	0.90	2.7×10^{-3}	0.15
Dz51 [e]	$\leq7\times10^{-5}$	-	-	-	$\leq7\times10^{-5}$	-	-	-
Dz122 [f]	3.8×10^{-2}	0.80	$\leq7\times10^{-5}$	-	5.0×10^{-2}	0.82	5.6×10^{-4}	0.30
Dz163 [f]	1.3×10^{-2}	0.90	$\leq7\times10^{-5}$	-	1.3×10^{-2}	0.90	2.7×10^{-4}	0.30

Table 2. Conditions: Tris-HCl buffer, pH 7.5, 10 mM Mg^{2+}, 37°C. Cleavage was performed under single-turnover conditions, using 10 nM substrate and 100 nM catalyst (DNAzyme or LNAzyme). In experiments with the 19-mer substrates, the 100 nM concentration was saturating for the LNAzymes (except for Lz51) but slightly subsaturating for the DNAzymes, as the observed reaction rates of Dz32, Dz122 and Dz163 increased by 1.5 to twofold when the DNAzyme concentration was raised to 1 mM. The kinetic time-courses were fit to Equation 1, where F_∞ indicates the final extent of cleavage. For each reaction, the reported k_{obs} values are the average of two to four independent experiments. k_{obs} values from individual experiments differed by no more than 20% from the mean, except for reactions of DNAzymes with the E6 RNA ('Protocol B' conditions) where the range of variation was within ±40% of the mean.

[a] When Lz51 was tested according to 'Protocol B', the kinetics of cleavage of the long RNA substrates were distinctly biphasic. The data were therefore fit to Equation 2, and the table reports the rate constants and the extent of cleavage for both phases. [b] To compare these data with those in Table 1, the reaction of Lz122 (100 nM) against the E6 transcript was also tested in the presence of 3 mM Mg^{2+}, yielding, $k_{obs}= 1.6\times10^{-2}$ min^{-1} and $F_\infty= 0.74$. [c] Lz163 cleaved its 19-mer substrate ~fourfold faster under the 'Protocol B' conditions than with 'Protocol A'. The behavior was reproducible but remains unexplained, since the minimal substrate of Lz163 is not expected to form any stable secondary structure. [d] When Dz32 was tested against the E6 transcript according to 'Protocol A', less than 5% of product was formed in ten hours. During this time, some unspecific degradation of the long mRNA also began to occur. Based on these data, and assuming an endpoint of 0.9, we estimated that cleavage by Dz32 would occur with a $k_{obs} \leq10^{-4}$ min^{-1}. [e] The reaction of 100 nM Dz51 with the 19-mer substrate was too slow to be reliably measured (no appreciable product formation in eight hours). By using a tenfold higher concentration of catalyst, we measured kinetics with an apparent endpoint of 0.60 and observed rate constants of $7x10^{-4}$ min^{-1} (with both protocols). [f] Upper limits for the rates of E6 cleavage were estimated as in d, except that less than 4% product was formed in ten hours.

Table 3
Comparing the performance of LNAzymes against the E6 and E6E7 transcripts.

	Protocol A (cleavage of pre-folded RNA)				Protocol B (cleavage after pre-annealing catalyst and substrate)			
	E6 mRNA		E6E7 mRNA		E6 mRNA		E6E7 mRNA	
	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞	k_{obs} (min^{-1})	F_∞
Lz32	4.4×10^{-2}	0.64	4.8×10^{-2}	0.83	4.5×10^{-2}	0.67	6.4×10^{-2}	0.87
Lz51	5.5×10^{-3}	0.44	7.0×10^{-3}	0.49	1^{st}: 5.2×10^{-3} 2^{nd}: 2.0×10^{-3}	F_1: 0.12 F_2: 0.64	1^{st}: 1.2×10^{-2} 2^{nd}: 3.7×10^{-3}	F_1: 0.16 F_2: 0.57
Lz122	2.5×10^{-2}	0.80	2.2×10^{-2}	0.74	4.1×10^{-2}	0.75	3×10^{-2}	73
Lz163	3.2×10^{-2}	0.85	3.3×10^{-2}	0.77	5.2×10^{-2}	0.86	5.3×10^{-2}	81

Table 3. The conditions are as described in the legend of Table 2 in the main text. Reactions were performed under single-turnover conditions (10 nM substrate + 100 nM catalyst) in 50 mM Tris-HCl buffer, pH 7.5, 10 mM Mg^{2+}, 37°C. Kinetic parameters (k_{obs} and F_∞) were obtained from exponential fits of the kinetic time-courses. When Lz51 was tested according to 'Protocol B', the kinetics of cleavage of E6 and E6E7 were distinctly biphasic. The data were therefore fit to a double exponential function, and the table reports the rate constants and the extent of cleavage for both phases.

Cleavage of the E6 mRNA under "pre-annealing" conditions

Could the performance gap between DNAzymes and LNAzymes be reduced by allowing the catalysts to encounter an unfolded target? To address this possibility, we measured cleavage of the E6 transcript using a distinct protocol (protocol B), where the LNAzyme (or DNAzyme) and its substrate were mixed together, heated at 80°C for two minutes and cooled at 37°C prior to starting the cleavage reaction. This protocol represents a stage between the cleavage of a fully folded mRNA and the cleavage of a minimal, unstructured substrate: the denaturation step at 80°C is expected to disrupt the local secondary and tertiary structures, thus allowing a direct access of the catalysts to their targets, even though binding of the catalysts (during cooling) would have to compete with the refolding process of the long RNA.

For the LNAzymes the differences between the data from the two protocols were mostly minor (**Table 2**). On the other hand, three of the all-DNA constructs, which were essentially inactive against the pre-folded long transcripts, showed in this case an appreciable ability to cleave the E6 RNA (**Table 2**, right), albeit with lower rates as compared to LNAzymes and with a modest overall yield.

Discussion

We have incorporated four or five LNA monomers in the arms of the 8-17 deoxyribozyme and shown that such a modification substantially improves the deoxyribozyme performances under a number of conditions. This result parallels those reported in previous studies on the 10-23 motif (Vester, 2002; Fahmy, 2004; Schubert, 2004; Fluiter, 2005; Vester, 2006), demonstrating that the inclusion of LNA monomers in the target-binding arms is advantageous for deoxyribozymes in general. More fundamentally, our work has addressed the mechanistic basis of the advantages conferred by LNAs, providing a detailed rationale for the effects reported by us and by other authors.

LNAzyme advantages in the cleavage of short, unstructured substrates

The kinetic data obtained with short RNA substrates indicate that deployment of LNA monomers in the 8-17 arms (at least, at positions not immediately adjacent to the central 'core') does not have major effects on the intrinsic rate of cleavage, described by k_2 in **Figure 1B**. This conclusion stems from the detailed comparison between Dz122 and Lz122 (**Table 1**) as well as from data on the Dz32/Lz32 and Dz163/Lz163 couples (see legend of **Table 2**). The point is not trivial, as previous studies suggested that alterations in the helices formed by the 8-17 arms could substantially affect catalysis (Li, 2000; Liu, 2004).

On the other hand, the comparison between Dz122 and Lz122 confirms that LNA monomers enhance binding of the deoxyribozyme to its target, raising the thermodynamic stability of the enzyme-substrate complex in consistency with published studies on short LNA-containing duplexes (Singh, 1998; Singh, 1998, Bondensgaard, 2000). In principle, this effect could partially arise from an improvement in hybridization kinetics, because the association rate constant (k_1) for the unmodified 8-17 is about 10-fold lower the constants typical for the annealing of short complementary oligonucleotides (Bonaccio, 2004; Santoro, 1998; Joyce, 2001).

However, k_1 values for Lz122 and Dz122 were very similar, in agreement with previous indications that LNAs do not accelerate substantially the hybridization between short, unstructured oligos (Christensen, 2001). Thus, the LNA-mediated increase in binding affinity can be attributed almost entirely to slowed-down kinetics of substrate release (**Table 1**).

The higher substrate affinity of the LNAzymes means that they can saturate their substrates and reach maximum catalytic efficiency at much lower concentrations than DNAzymes having the same sequence. For example, **Figure 6A** shows that, in reactions carried out at 1 nM enzyme, Lz122 is about tenfold more efficient than Dz122 in cleaving a minimal substrate. This gap is predicted to become even wider if lower concentrations of catalyst were used: the maximum difference in k_{obs} (attainable at enzyme concentrations well below the K_S of Lz122) is predicted to be ~30-fold, directly mirroring the difference in affinity between LNAzyme and DNAzyme (**Figure 6A**).

It is very likely that the 10-23 LNAzyme constructs described in previous studies (Vester, 2002; Schubert, 2003; Fahmy, 2004; Schubert, 2004; Fluiter, 2005; Vester, 2006; Jakobsen, 2007) behave similar to Lz122. The lower concentration requirements may explain, for example, earlier data showing that the activity of 10-23 against a short RNA substrate was improved by including LNA monomers within the deoxyribozyme arms (Vester, 2002). In fact those data were collected at nanomolar concentrations of catalyst – that is, under conditions that presumably did not allow efficient substrate binding by the all-DNA construct, but permitted a stable binding and cleavage by the LNAzyme.

LNAzyme advantages in the cleavage of long, structured substrates

In the presence of long substrates, the disparity in performance between DNAzymes and LNAzymes is impressive, and does not seem to simply reflect the diverse binding affinities for the two types of catalysts. In particular, we just noted that the difference in activity between Dz122 and Lz122, predicted from experiments with a minimal substrate and attributable to the relative binding affinities, should be at most ~30-fold. The actual difference observed against the long transcripts, however, greatly exceeds this value (**Table 2**, left). More fundamentally, the differences in substrate binding affinities (K_S) or in substrate dissociation rates (k_{-1}) would not explain why the performance gap between DNAzymes and LNAzymes depends on the reaction protocol, i.e. why cleavage of the E6 RNA by the DNAzymes (but not LNAzymes) is improved by pre-annealing the catalysts to their targets (**Table 2**).

When DNAzymes and LNAzymes are reacted with the fully folded E6 transcript (i.e., under Protocol A conditions), the local RNA structure represents a kinetic barrier to binding of the catalysts, which is expected to slow down the

association process; under this condition, LNAzymes performed well, whereas the activity of DNAzymes was hardly measurable. The heating and cooling step in protocol B allows a direct access of the catalysts to their target sequences, removing at least in part the initial kinetic barrier to association; in this case, Dz32, Dz122 and Dz163 showed a substantially increased activity (**Table 2**). Hence, our data suggest that annealing to a folded substrate is *kinetically* much more difficult for DNAzymes than for LNAzymes. Put another way, the data in table 2 suggest that LNAzymes bind to structured targets with a faster association rate (k_1) as compared to DNAzymes.

This conclusion is not completely unexpected. Previous studies on 10-23 constructs had implicitly suggested that LNA facilitates the kinetics of deoxyribozyme annealing to structured RNA (Vester, 2002; Schubert, 2004; Vester, 2006). Moreover, in contrast with earlier results obtained with short, *unstructured* oligos (Christensen, 2001), Ormond et al. recently showed that the hybridization between *structured* oligonucleotides (i.e., oligos forming an intramolecular stem-loop) could be accelerated by including a few LNA residues in one of the two complementary strands (Ormond, 2006).

A model for LNAzymes reacting with structured substrates

LNA did not greatly affect the kinetics of 8-17 binding to a short substrate, so why should things be different with long, folded RNAs? Binding of a DNAzyme to a structured target is almost certainly a more complex phenomenon than annealing to a minimal substrate, and presumably occurs through the formation of intermediates in which the hybridization is only partially complete (Reynaldo, 2000; Walton, 2002). For example, a simple two-step model can be envisaged, whereby the enzyme initially base-pairs to a stretch of just a few accessible residues (nucleation), after which the hybridization proceeds to completion, leading to disruption of the local RNA secondary structure (**Figure 8**).

Analogous models have been suggested before in other systems. For example, while this paper was in preparation, a model very similar to the one in Figure 5 was proposed to explain the binding of microRNAs to their target sequences within structured mRNAs; the model was also shown to account accurately for a series of *in vivo* results (Long, 2007). MicroRNAs and deoxyribozymes are both short oligonucleotides with an extended, but not perfect, complementarity to their targets,

thus it seems reasonable that their binding mechanisms may share several common features.

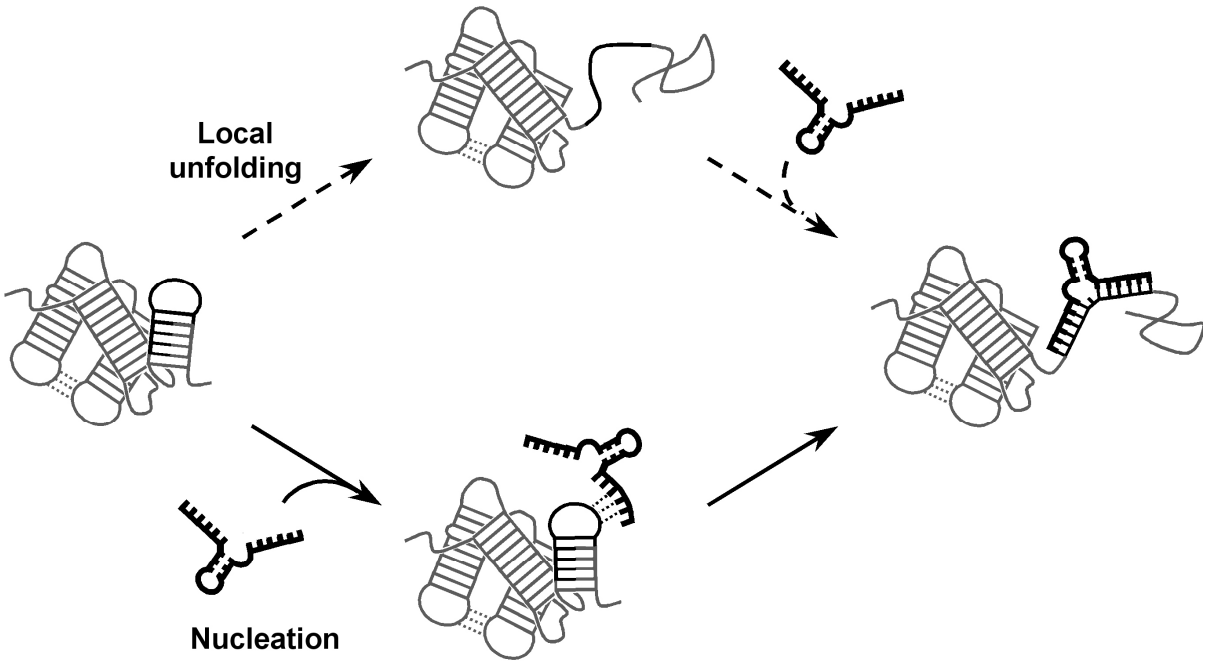


Figure 8. A schematic model for the binding of DNAzymes and LNAzymes to target sequences embedded in large, structured RNAs. In principle, the catalyst may bind to its target when the target sequence becomes completely accessible, thanks to a local unfolding event (e.g., the occasional fraying of a double helix; upper pathway). Binding to a fully accessible target, however, is presumably very rare in the context of a large mRNA, at physiological temperatures. More often, complete hybridization of the catalyst may entail formation of binding intermediates, which precede and favor the disruption of the local mRNA structure. For example, the enzyme might initially form interactions with a few unpaired nucleotides on the target ('nucleation'; lower pathway). Such interactions would be particularly stable for LNA-armed catalysts, thereby promoting a more efficient disruption of the local structure and a faster completion of binding. Note that other types of intermediates (e.g., triple-stranded species) might be formed during the initial nucleation events: whatever the nature of these transient intermediates, their stabilization might ease the overall binding process.

The two-step model offers an explanation as to why LNAzymes can bind faster than DNAzymes to structured targets. In fact, normal DNAzymes would form relatively weak interactions in the initial nucleation complex, so that the catalyst would generally fall off before completing the hybridization process. For the LNAzymes, however, nucleation would yield a more stable binding intermediate, providing a bridgehead for the invasion of the nearby structure and therefore promoting a faster formation of the catalyst-substrate complex. Once formed, this complex would be more stable than the corresponding DNAzyme-substrate complex

(due to the intrinsically lower k_{-1}), thus reducing the chances that the catalyst may dissociate before cleaving.

While the model in **Figure 8** certainly needs to be further refined and tested, it represents a useful starting point for future mechanistic analyses. The model has implications not just for LNAzymes, but also, more generally, for oligonucleotides containing residues that form especially strong basepairs with RNA – be they LNA, 2'-O-methyl nucleotides or other monomers. The model predicts that, broadly speaking, these modified oligonucleotides will bind to structured RNA at a faster rate as compared to unmodified oligos. The actual extent of acceleration would presumably depend on factors such as the number and location of the modified monomers within the oligonucleotide.

Multiple turnover and specificity issues

The data obtained in this work demonstrate the advantages of introducing LNA monomers in the deoxyribozyme arms. In principle, however, the use of LNAzymes may be disadvantageous under some conditions.

In particular, LNA-armed deoxyribozymes are expected to interact strongly with the cleavage products - a feature that may decrease the reaction rate under multiple turnover conditions. This problem, however, should arise only when the deoxyribozyme arms contain a high number of LNA monomers. Indeed, recent studies have shown that LNAzymes containing just two LNA monomers on each arm (similar to Dz122) function usually better than plain DNAzymes under multiple-turnover conditions (Vester, 2006, Jakobsen, 2007).

The possibility of off-target binding is also an important concern. In fact, LNA monomers could favor annealing of a deoxyribozyme not only to the intended target, but also to sequences that differ from the target by just a few nucleotides. Nevertheless data in the literature suggest that mismatches involving LNA monomers are not more stable than normal mismatches (You, 2006) and that, in general, LNA improves the single-nucleotide mismatch discrimination with respect to that obtained with DNA (You, 2006; Mouritzen, 2003). We also note that the LNAzymes used in this study showed no sign of off-target cleavage within the E6 mRNA (**Figure 7A**) or the E6E7 mRNA (not show).

Conclusions

We have analyzed in detail the kinetic and thermodynamic properties of LNA-containing deoxyribozymes, highlighting the mechanistic differences between these catalysts and their all-DNA counterparts. Overall, the results of this work emphasize that the advantageous features of LNAzymes become fully manifest when performing reactions at low catalyst concentrations, at low magnesium concentrations and against large, structured RNAs. These are exactly the conditions that are expected to occur in cells, helping to explain why LNA-armed deoxyribozymes have proven substantially superior to conventional DNAzymes for the knockdown of specific mRNAs *in vivo* (Fahmy, 2004; Jakobsen,2007).

Acknowledgements

I would like to thank M. Tommasino (IARC, Lyon) for the plasmids carrying the E6 and E6E7 genes. I would also like to thank Lykke H. Hansen for excellent technical help and Holger Døssing for the computational analysis of RNA secondary structures. SD was supported by a Marie Curie Early Stage Research Training Fellowship of the European Community's Sixth Framework Programme under Contract Number MEST-CT-2004-504018.

References

Bonaccio, M., Credali, A., and Peracchi, A. (2004) *Nucleic Acids Res.* 32, 916-925

Bondensgaard, K., Petersen, M., Singh, S. K., Rajwanshi, V. K., Kumar, R., Wengel, J., and Jacobsen, J. P. (2000) *Chemistry* 6, 2687-2695

Cairns, M. J., Hopkins, T. M., Witherington, C., Wang, L., and Sun, L. Q. (1999) *Nature Biotechnol.* 17, 480-486

Cairns, M. J., Saravolac, E. G., and Sun, L. Q. (2002) *Curr. Drug Targets* 3, 269-279

Cairns, M. J., and Sun, L. Q. (2004) *Methods Mol. Biol.* 252, 267-277

Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) *Biopolymers* 9, 1059-1077

Christensen, U., Jacobsen, N., Rajwanshi, V. K., Wengel, J., and Koch, T. (2001) *Biochem. J.* 354, 481-484

Dass, C. R. (2004) *Trends Pharm. Sci.* 25, 395-397

Dass, C. R. (2006) *Drug. Dev. Ind. Pharm.* 32, 1-5

Fahmy, R. G., and Khachigian, L. M. (2004) *Nucleic Acids Res.* 32, 2281-2285

Ferrari, D., and Peracchi, A. (2002) *Nucleic Acids Res.* 30, e112

Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, Freeman & Co., New York, NY

Fluiter, K., Frieden, M., Vreijling, J., Koch, T., and Baas, F. (2005) *Oligonucleotides* 15, 246-254

Jakobsen, M. R., Haasnoot, J., Wengel, J., Berkhout, B., and Kjems, J. (2007) *Retrovirology* 4, 29

Joyce, G. F. (2001) *Methods Enzymol.* 341, 503-517

Kurreck, J., Bieber, B., Jahnel, R., and Erdmann, V. A. (2002) *J. Biol. Chem.* 277, 7099-7107.

Kaur, H., Arora, A., Wengel, J., and Maiti, S. (2006) *Biochemistry* 45, 7347-7355

Schubert, S., Gul, D. C., Grunert, H. P., Zeichhardt, H., Erdmann, V. A., and Kurreck, J. (2003) *Nucleic Acids Res.* 31, 5982-5992

Schubert, S., Furste, J. P., Werk, D., Grunert, H. P., Zeichhardt, H., Erdmann, V. A., and Kurreck, J. (2004) *J. Mol. Biol.* 339, 355-363

Vester, B., Hansen, L. H., Lundberg, L. B., Babu, R., Sorensen, M. D., Wengel, J., and Douthwaite, S. (2006) *BMC Mol. Biol.* 7, doi:10.1186/1471-2199-1187-1119

Li, J., Zheng, W., Kwon, A. H., and Lu, Y. (2000) *Nucleic Acids Res.* 28, 481-488.

Liu, Y., and Sen, D. (2004) *J. Mol. Biol.* 341, 887-892

Long, D., Lee, R., Williams, P., Chan, C. Y., Ambros, V., and Ding, Y. (2007) *Nat. Struct. Mol. Biol.* 14, 287-294

McTigue, P. M., Peterson, R. J., and Kahn, J. D. (2004) *Biochemistry* 43, 5388-5405

Mouritzen, P., Nielsen, A. T., Pfundheller, H. M., Choleva, Y., Kongsbak, L., and Moller, S. (2003) *Expert Rev. Mol. Diagn.* 3, 27-38

Nakano, S., Fujimoto, M., Hara, H., and Sugimoto, N. (1999) *Nucleic Acids Res.* 27, 2957-2965.

Ormond, T. K., Spear, D., Stoll, J., Mackey, M. A., and St John, P. M. (2006) *J. Biomol. Struct. Dyn.* 24, 171-182

Pan, W. H., Xin, P., Bui, V., and Clawson, G. A. (2003) *Mol. Ther.* 7, 129-139

Peracchi, A., Bonaccio, M., and Clerici, M. (2005) *J. Mol. Biol.* 352, 783-794

Reynaldo, L. P., Vologodskii, A. V., Neri, B. P., and Lyamichev, V. I. (2000) *J. Mol. Biol.* 297, 511-520

Santoro, S. W., and Joyce, G. F. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 4262-4266

Santoro, S. W., and Joyce, G. F. (1998) *Biochemistry* 37, 13330-13342

Silverman, S. K. (2005) *Nucleic Acids Res.* 33, 6151-6163

Singh, S. K., Nielsen, P., Koshkin, A. A., and Wengel, J. (1998) *Chem. Commun.*, 455-456

Singh, S. K., and Wengel, J. (1998) *Chem. Commun.*, 1247-1248

Venturini, F., Braspenning, J., Homann, M., Gissmann, L., and Sczakiel, G. (1999) *Nucleic Acids Res.* 27, 1585-1592

Vester, B., Lundberg, L. B., Sorensen, M. D., Babu, B. R., Douthwaite, S., and Wengel, J. (2002) *J. Am. Chem. Soc.* 124, 13682-13683.

Vester, B., and Wengel, J. (2004) *Biochemistry* 43, 13233-13241

Walton, S. P., Stephanopoulos, G. N., Yarmush, M. L., and Roth, C. M. (2002) *Biophys. J.* 82, 366-377

Wengel, J. P. (1999) *Accounts Chem. Res.* 32, 301-310

Zuker, M. (2003) *Nucleic Acids Res.* 31, 3406-3415

You, Y., Moreira, B. G., Behlke, M. A., and Owczarzy, R. (2006) *Nucleic Acids Res.* 34, e60

List of Pubblications

1. Donini S, Percudani R, Credali A, Montanini B, Sartori A, Peracchi A (2006). A threonine synthase homolog from a mammalian genome. *Biochem. Biophys. Res. Commun.* **350** 922-928.

2. Donini S, Clerici M, Wengel J, Vester B, Peracchi A. (2007). The advantages of being locked: assessing the cleavage of short and long RNAs by locked nucleic acid-containing 8-17 deoxyribozymes. *J. Biol. Chem.* **282** 35510-35518.

Ringraziamenti

Un grazie al mio supervisore, Alessio Peracchi. Grazie per la supervisione e gli stimoli che mi hai dato in questi anni, grazie per avermi dato la possibilità di lavorare in una realtà completamente diversa come quella dell'SDU……e per la correzione della tesi!!!

Grazie alla mia famiglia per il supporto in questi anni e per il "rifornimento" del fine settimana.

Un grazie ai compagni del lab, in particolare a Roby e Chiara i peggiori coinquilini del mondo.