Tyrosine metabolism: identification of a key residue in the acquisition of prephenate aminotransferase activity by 1β aspartate aminotransferase

Cécile Giustini, Matthieu Graindorge, David Cobessi, Serge Crouzy, Adeline Robin, Gilles Curien, Michel Matringe

Tyrosine metabolism: identification of a key residue in the acquisition of prephenate aminotransferase activity by 1β aspartate aminotransferase.

Cecile Giustini[1], Matthieu Graindorge[1], David Cobessi[2], Serge Crouzy[3], Adeline Robin[1], Gilles Curien[1], Michel Matringe[1*]

[1]Univ. Grenoble Alpes, INRA, CNRS, CEA, BIG-LPCV, 38000 Grenoble France.
[2]Univ. Grenoble Alpes, CNRS, CEA, IBS, 38000 Grenoble France.
[3]Univ. Grenoble Alpes, CEA, CNRS, BIG-LCBM, 38000 Grenoble France

Running title: *Signature of prephenate competent 1β AAT*

To whom correspondence should be addressed: Michel Matringe, Laboratoire de Physiologie Cellulaire et Végétale, CEA, BIG, 17 rue des Martyrs, Grenoble 38054, France. Tel.: 33 4 38 784 184; Fax: 33 4 38 785 091; E-mail: michel.matringe@cea.fr

Keywords: tyrosine metabolism, prephenate, arogenate route, 1β aspartate aminotransferase, structure function, molecular modeling

Abreviation: AAT Aspartate aminotransferase, PAT Prephenate aminotransferase

ABSTRACT

Alternative routes for the post-chorismate branch of the biosynthetic pathway leading to tyrosine exist, the 4-hydroxyphenylpyruvate or the arogenate route. The arogenate route involves the transamination of prephenate into arogenate. In a previous study, we found that, depending on the microorganisms possessing the arogenate route, three different aminotransferases evolved to perform prephenate transamination *i.e.* 1β aspartate aminotransferase (1β AAT), *N*-succinyl-L,L-diaminopimelate aminotransferase, and branched-chain aminotransferase. The present work aimed at identifying molecular determinant(s) of 1β AAT prephenate aminotransferase (PAT) activity. To that purpose, we conducted X-ray crystal structure analysis of two PAT competent 1β AAT from *Arabidopsis thaliana* and *Rhizobium meliloti*, and one PAT incompetent 1β AAT from *Rhizobium meliloti*. This structural analysis supported by site directed mutagenesis, modelling and molecular dynamics simulations allowed us to identify a molecular determinant of PAT activity in the flexible N-terminal loop of 1β AAT. Our data reveal that a Lys/Arg/Gln residue in position 12 in the sequence (numbering according to *Thermus thermophilus* 1β AAT), present only in PAT competent enzymes, could interact with the 4-hydroxyl group of the prephenate, substrate and thus may have a central role in the acquisition of PAT activity by 1β AAT.

The aromatic amino acid biosynthesis pathway occurs in microorganisms, fungi and plants, and is not found in animals. This pathway links the metabolism of carbohydrates to the biosynthesis of aromatic amino acids involved in protein synthesis, and leads to the synthesis of a large diversity of aromatic secondary metabolites [1-3]. The aromatic amino acid biosynthesis pathway plays a central role for the growth of these organisms and their interaction with environment. Phenylalanine and tyrosine also represent an important source of high-value aromatic natural compounds for industry, and human health.

Two alternative routes leading from chorismate to tyrosine and phenylalanine exist: the 4-hydroxyphenylpyruvate route, and the arogenate route (Figure 1). All photosynthetic organisms and many microorganisms synthesize L-tyrosine *via* the arogenate route [4-7]. In addition, most viridiplantae and some microorganism also synthesize L-phenylalanine *via* the arogenate route [8]. In this case, a prephenate aminotransferase transforms prephenate into arogenate, and an arogenate dehydrogenase (EC 1.3.1.43) or an arogenate dehydratase (EC 4.2.1.91) decarboxylates arogenate into L-tyrosine or L-phenylalanine, respectively (Figure 1) [7,8]. Alternatively, some microorganisms and fungi such as *Escherichia coli*, or yeast, first transform prephenate into 4-hydroxyphenylpyruvate by a prephenate dehydrogenase (EC 1.3.1.13) or into penylpuruvate by a prephenate dehydratase (EC 4.2.1.51) [9], and then both are transaminate into L-tyrosine or L-phenylalanine, respectively.

Prephenate aminotransferase (PAT) activity was first identified in plants [10,11]. In *Arabidopsis thaliana*, a specific chloroplastic aspartate aminotransferase is in charge of PAT activity (At2g22250 gene, EC 2.6.1.1, EC 2.6.1.78, and EC 2.6.1.79) [10,11]. This prephenate aminotransferase belongs to aspartate aminotransferase (AAT) subgroup 1β [12]. In this subgroup, the dicarboxylic substrate interacts with a lysine and an arginine (K101 and R361, numbering according to *Thermus thermophilus* 1β AAT enzyme: Q56232 abbreviated as Tth-PAT) instead of two arginines in the 1α AAT subgroup. In a previous study, we showed that the PAT competent 1β AAT from *A. thaliana* (Ath-PAT) catalyzes the transamination of oxaloacetate into aspartate as efficiently as prephenate into arogenate [10].

Historically in microorganisms, arogenate route has been detected in cyanobacteria, actinobacteria, some α- and β-proteobacteria by the group of Jensen [4-6,13]. In a previous study [14] we reported that, in these arogenate competent microorganisms, PAT activity is carried out by three different aminotransferases, which present two different fold-types. Furthermore, we showed that the capacity to transaminate prephenate is not a general catalytic property of the three classes of aminotransferases, since other members were devoid of detectable PAT activity [14]. These results revealed that the arogenate route emerged independently several times during evolution. To go further, structure-function studies are needed to identify molecular signature(s) of the competent prephenate/arogenate aminotransferases, in order to investigate their presence in genomic databases.

In the present study, by combining X-ray crystal structure analysis of two PAT competent 1β AAT from *Arabidopsis thaliana* and *Rhizobium meliloti*, and one PAT incompetent 1β AAT from *Rhizobium meliloti*, site directed mutagenesis, modelling and molecular dynamics simulations, we show that K/R/Q12 residue (numbering according to *T. thermophilus* 1β AAT) is a signature of the PAT function of 1β AAT. It is present in the N-terminal flexible loop only in PAT competent 1β-AAT and our data strongly support its possible role in stabilizing prephenate by interacting with its 4-hydroxy group.

Results and discussion

Resolution of 3D structures of PAT competent and PAT incompetent 1β AAT.

In our previous study [14], of the two 1β AATs from *Rizobium meliloti* tested, one (Q02635) showed a good prephenate aminotransferase activity while the other (P58350) had no detectable prephenate aminotransferase activity under our test conditions.

We thus decided to identify molecular determinant(s) involved in the acquisition of prephenate aminotransferase function by the 1β AAT through a comparative X ray structural study of PAT competent and PAT incompetent 1β AAT. To do this, we solved the three-dimensional structures of PAT competent from *Arabidopsis thaliana* (Ath_Q9SIE1, Ath-PAT) and from *Rizobium meliloti* (Rme_Q02635, Rme-PAT), and the PAT incompetent 1β AAT from *Rizobium meliloti* (Rme_P58350, Rme-AAT).

Overall structures

The 3D structures of PAT competent 1β AAT Ath-PAT, and Rme-PAT and PAT incompetent 1β AAT Rme-AAT show homodimeric proteins, covalently bound to pyridoxal phosphate (PLP) and sharing a common fold (Figure 2). The monomers are related by a non-crystallographic 2-fold symmetry axis. Residues in or preceding the first α-helix are observed in different positions when monomers of Ath-PAT are compared, suggesting flexibility. A citrate molecule is bound to the catalytic site of 3 monomers of Ath-PAT (two dimers are in the asymmetric unit). It is stabilized by interacting with Lys122, Trp146, Asn196, Arg398 side chains, the Gly60 NH group and water molecules. In Rme-PAT, pyridoxal phosphate cofactor is not covalently bound to the protein probably as a consequence of radiation damages.

Structure comparisons

The Cα coordinates rms deviations calculated between all the monomers are low indicating that the structures are globally very similar. The largest difference is observed when Rme-AAT is compared with Ath-PAT (1.56 Å) or Tth-PAT (PDB code 1BKG) (1.63 Å). Dimer comparison after superimposing monomers shows moderate differences in the positioning of monomers. The catalytic site is large and solvent accessible and residues from the two monomers participate to the catalytic pockets. Most of the residues observed in the Ath-PAT catalytic site are conserved in Rme-PAT, Rme-AAT and Tth-PAT. From structure comparison with Tth-PAT, residues from the first α-helix (Lys33(A), Pro34(A), Ser35(A), Lys36(A), Thr37(A), Met38(A) in Ath-PAT [15,16]) are also involved in the closing of the catalytic site

upon ligand binding. Lys259 of Ath-PAT is covalently bound to PLP through an aldimine link. The interactions observed between maleate and residues of the catalytic site in Tth-PAT could also take place in Ath-PAT. Maleate observed in the Tth-PAT catalytic site is superimposable to citrate observed in Ath-PAT, but the position of the first α-helix is different since the Ath-PAT catalytic site is in open conformation.

3D X ray structures of PAT competent 1β AAT Ath-PAT and Rme-PAT, and PAT incompetent 1β AAT Rme-AAT, present a very similar organization. Comparison of these three 3D structures did not reveal a clear molecular determinant of PAT activity, most certainly because they are in an open conformation. Indeed, during the catalytic reaction, the keto acid substrate (oxaloacetate or prephenate) binds to a PMP-enzyme complex and a short N-terminal flexible loop move to close the active site [15,16]. Only a closed conformation of prephenate-PMP-enzyme complex is relevant for structural comparison of PAT competent and incompetent 1β AAT. Unfortunately, all our attempts to obtain either a PMP-prephenate- or PMP-maleate-enzyme complex of Rme-PAT or Ath-PAT failed.

However, the observation that the 1β AAT from *Thermus thermophilus* is PAT competent ([17], this study) allowed us to use its 3D structure in a closed conformation, in complex with PMP and maleate (PDB code 1BKG), resolved by Nakai *et al*. [15], to model the binding of prephenate in the active site.

Identification of a potential molecular determinant of PAT capacity of 1β AAT

Taking advantage of molecular elucidation of enzyme/ligand interactions in PAT competent 1β AAT from *T. thermophilus* by Nakai et al [15], we manually replaced the substrate analogue maleate with prephenate in the active site of the 3D structure 1BKG. Prephenate differs from oxaloacetate and ketoglutarate, the two ketoacid substrates of aspartate aminotransferase, by the presence of its 2,5-cyclohexadiene-4-hydroxy group (Figure 3A). The idea is therefore to identify potential residue(s) present only in the PAT competent 1β AAT that could interact with the 2,5-cyclohexadiene-4-hydroxy group of prephenate. We placed the two prephenate carboxylate groups in interaction with Lys101 and Arg361 while its keto group points to the PMP NH2 group and the catalytic lysine, Lys234 (Figure 3B). In this position, the 2,5- cyclohexadiene-4-hydroxy group of the prephenate substrate points toward the entrance of the active site, its 4-hydroxyl facing the flexible N-terminal loop $_{10}$AMKPSA$_{15}$, which is known to move and close the active site while the substrate binds to the active site [15,16]. In that context, it is interesting to note that the alignment of the N-terminal sequences of already characterized PAT competent 1β AAT (*Arabidopsis thaliana* [10], *Sorghum bicolor* [18], *R. meliloti*, *Rhodobacter sphaeroides* [14], *T. thermophilus* [17] (this study) and *Chlorobium tepidum* [17]) with those of the already characterized PAT incompetent 1β AAT (*R. meliloti, Streptomyces avermitilis* and *Synechocystis* [14]) (Figure 4), reveals that a lysine residue belonging to this flexible loop, Lys12 (numbering according to *T. thermophilus* 1β AAT), is present only in the PAT competent 1β AAT or is replaced by an arginine in all Poaceae like *S. bicolor* or *Zea mays* or a glutamine in some α Proteobacteria like *Rhodobacter sphaeroides* and in Chlorobi like *Chlorobium tepidum*. Each of these three amino acids has the ability to stabilize prephenate in the active site by interacting with its 4-hydroxyl group. In contrast, in PAT incompetent 1β AAT, Lys12 is replaced by a glycine, a valine or a serine. None of the later three can play the same role.

However, in the ternary maleate-PMP-enzyme complex (1BKG), the Lys12 side chain points outside the active site [15] (Figure 3). We hypothesize that in the presence of prephenate, the Lys12 side chain rotates to interact with the prephenate cyclohexadiene-4-hydroxy group during closure of the active site, thereby stabilizing the prephenate within the active site.

Lys12 plays a specific role in prephenate binding by PAT competent 1β AAT from *R. meliloti* and *T. thermophilus*

The potential interaction of Lys12 with the prephenate 4-hydroxyl group was first addressed by using a site directed mutagenesis approach performed on PAT competent 1β AAT from *R. meliloti* and *T. thermophilus*. We replaced the lysine of the $_{10}$AMKPSA$_{15}$ motif by a glycine. Steady-state kinetic properties of wild-type and K12G mutants were determined with a fixed concentration (25mM) of Glutamate as amino donor and variable concentrations of either prephenate or oxaloacetate. Modification of the apparent Michaelis-Menten constant (K_M) will reflect a modification in the affinity of the 1β AAT K12G mutants for the substrate whereas modification of the apparent catalytic constant (k_{cat}) will reflect an impact of the

mutation on the reaction rate. If our assumption was correct we anticipate that the mutation will have a significant impact on the affinity for prephenate, much less on the affinity for oxaloacetate, and will have minor impact on the reaction rate of both oxaloacetate and prephenate aminotransferase activity. Apparent K_M for prephenate was found 7 time higher in the K12G mutant of PAT competent 1β AAT from *R. meliloti* (Q02635) (K12G: 692±143 µM versus WT: 106±12µM). In the other hand, apparent K_M for oxaloacetate was only weakly affected in K12G mutant (Table 3). Furthermore, the apparent catalytic constants of the two aminotransferase reactions (k_{cat}) were not affected.

The effect of the K12G mutation was also tested on the PAT activity of the *T. thermophilus* 1β AAT (Q56232). The K12G mutation gave similar results with a 10 time increase of the apparent K_M for prephenate (K12G: 2332±204 µM versus WT: 150± 8.7 µM) and no effect on the apparent K_M for oxaloacetate (WT: 25.3±1.8 µM versus K12G: 26.4±2.3 µM). It should be noted that in our conditions, the *T. thermophilus* 1β AAT apparent catalytic activity is slightly affected by the K12G mutation (k_{cat} oxaloacetate WT: 32.2±4.6 s^{-1} versus K12G: 12.3±3.7 s^{-1}). Prephenate aminotransferase apparent catalytic activity seems less affected by the mutation (k_{cat} for prephenate WT: 7.7±1.6 s^{-1} versus K12G: 4.8 ± 2.1 s^{-1}), although the strong degradation of affinity for prephenate of the K12G mutant makes the determination of its kinetic parameters less accurate. Steady state kinetics analysis confirms thus that in both enzymes K12G mutation induces a specific decrease in affinity for prephenate with limited effect on their prephenate aminotransferase reaction rate.

In our previous study [14], the two incompetent PATs of *R. meliloti* (P58350), or *Synechocystis* (Q55128) which respectively have a G or a V instead of K12, were found to have no PAT activity at prephenate concentrations up to 2.5 mM, by our coupled test. However, in the present study, the two K12G mutants maintain prephenate aminotransferase activity at high prephenate concentrations since the K12G mutation does not affect their reaction rate.

In their study, Dornfield et al [17] propose that two highly conserved residues among 1β AAT, T84 and K169, (numbering according to *A. thaliana* 1β AAT, T16 and L101 according to *T. thermophilus* 1β AAT) (Figure 3 and 4), are important residues for substrate recognition in PAT competent 1β AAT. Furthermore, during the final stage of writing our manuscript, a new structural study on Arabidopsis PAT competent 1β AAT (At2g22250) from the same group was published in Plant Journal [16]. In this study, the authors proposed that, in addition to T84 and K169, E108 (numbering according to *A. thaliana* 1β AAT, E40 according to *T. thermophilus* 1β AAT) (Figure 3 and 4) plays an important role in the acquisition of PAT competencies by 1β AAT. However, unlike K12, these residues are conserved among all (K169, and E108) or almost all (T84) 1β AAT, regardless their ability to catalyze the transamination of prephenate, or not [14,17]. Therefore, although these residues may have contributed to the evolution of prephenate aminotransferase activity within class 1β AAT and not class 1α AAT, they can hardly be considered to have a specific role in the prephenate aminotransferase activity of 1β AAT. On the contrary, the presence of K/R/Q12 only in PAT competent 1β AAT, as well as the strong and selective impact of the K12G mutation on the binding affinity for prephenate, strongly support the specific role played by Lys12 in the binding of prephenate, most likely by its stabilization at the active site through interaction with its 4-hydroxyl group.

A clear support to the specific role played by Lys12 in the binding affinity for prephenate was afforded by Nature through the case of *Synechocystis* Q55128 and *Synechococcus* Q3AK33 1β AAT. As already mentioned *Synechocystis* Q55128 1β AAT have a V in place of a K, and was found inactive in our coupled assay [14]. However, in their study Dornfield et al [17] were able to detect traces of arogenate by HPLC in their end point test [17]. They reported a weak activity with an apparent K_M for prephenate in the range of 2600 µM (in the range of the apparent K_M for prephenate of the *T. thermophilus* 1β AAT K12G mutant). Blast search revealed that the Cyanobacteria *Synechococcus* Q3AK33 1β AAT, which is a close relative to *Synechocystis* Q55128 1β AAT, have a K in place of a V (Figure 5A). In order to compare the catalytic properties of these two 1β AAT, we have cloned, expressed in *E. coli* and partially purified the recombinant Q3AK33 1β AAT (Figure 5B). In our conditions, in contrast with *Synechocystis* Q55128 [14], a weak prephenate activity could be detected with the recombinant *Synechococcus* Q3AK33 1β AAT by our coupled assay. Kinetic characterization of prephenate aminotransferase activity of recombinant Q3AK33 (Figure 5C) reveals that its apparent K_M for prephenate was found in the range of 207±23 µM µM versus 2600 µM for *Synechocystis* Q55128 1β AAT [17], nicely confirming that the presence of K12 strongly

improves binding affinity for prephenate. Our data also indicate that other residues that remains to identify are responsible for the high reaction rate exhibited by all tested 1β AAT-PAT from higher plants [10 11] and by some 1β AAT from bacteria like *Rhizobium meliloti* (Q02635) [14] or *Chlorobium tepidum* (Q8KDS8) [17].

Structural evidence for Lys12 involvement in prephenate stabilization within the active site

We wanted to confirm, at the structural level, the specific role of Lys12 pointed out by the structure comparisons of PAT competent and incompetent 1β AAT, and by kinetics analysis of K12G mutants. Unfortunately, as already mentioned, all our attempts to obtain either a PMP-prephenate-enzyme complex of PAT competent 1β AAT from *A. thaliana*, *R. meliloti* or *T. thermophilus* failed, and at that time the only available 3D structure of a PAT competent 1β AAT in a closed conformation was the PMP-maleate-enzyme complex of 1β AAT from *T. thermophilus*. (PDB code 1BKG), previously solved by Nakai *et al*. [15]. We thus used this 3D structure as a starting model to perform structural analyses *via* modeling and molecular dynamics simulations of the PMP-prephenate-enzyme complex to support the interaction between Lys12 and prephenate suggested by our kinetic experiments.

After model building, introduction of sodium counter-ions and solvation in a periodic box of TIP3P water molecules (structure *1BKG_Preph_Mod*, see methods), a short equilibration MD simulation of the PAT dimer in the presence of PMP and prephenate was run at 300 K for 20 ns leading to structure *1BKG_Preph_MD* (See methods). Starting from a structure with low energy and low Lys12 NZ / prephenate O1 oxygen distance (dK-Prf) of 3.2Å from this equilibration, we have run a QM/MM optimization followed by a calculation of the charges adequate to fit the calculated electrostatic potential around the ligands (PMP and prephenate). In addition to prephenate, PMP and Lys12, the QM part consisted of side chains of Lys101, Trp125, Ser127, Lys234, Arg242, Tyr322, Arg321 from chain A and Tyr64 from chain B. 14 water molecules were also included in this quantum part for a total of 191 QM atoms. All MM atoms belonging to residues within a distance of 3 Å of the QM part were allowed to relax during geometry optimizations, and the remaining structure was kept frozen.

The optimized structure of the catalytic site (see materials and methods for details) is shown in Figure 6 and 7. Stabilizing interactions through hydrogen bonds exist between Lys12, Lys101, Trp125, Arg361, Tyr322 and prephenate, between prephenate and PMP and between PMP and Tyr64, Arg242, Tyr322 and Lys234.

The atomic partial charges calculated using the Merz-Kollman routine [20] in Gaussian were directly transferred to the protein structure file describing the MM system for prephenate, PMP and surrounding residues.

Molecular dynamics simulations

Using this modified structure file, 24 ns constant pressure and temperature CPT (Constant Pressure and Temperature) dynamics were run restarting from structure *1BKG_Preph_MD*. The nuclear Overhauser enhancement (NOE) type restraints used to obtain *1BKG_Preph_MD* (see materials and methods for details) were kept for 4 ns and then removed for the remaining 20 ns. This distance dK-Prf oscillates between 2.5 and 4.5 Å during the first 20 ns and increases to up to 14 Å when the restraints are released (See Figure 8). Interestingly, this distance drops back down to around 2.8 Å after 42 ns, meaning that without any restraint on the protein, the distance between Lys12 and prephenate is occasionally compatible with the reaction proposed in the experimental section.

Improving conformational sampling

The previous MD simulation confirms that the N-terminal part of the protein could undergo conformational transitions during substrate binding as shown by the structural study of Nakai *et al* [15], enabling Lys12 to be stabilized in the vicinity of the prephenate. Still, a method to explore more conformational space is to use Replica eXchanging Self-Guided Langevin Dynamics [21]. 16 replicas were created starting from structure *1BKG_Preph_Mod* and vacuum RXGLD simulations were run for 10 ns in the presence of restraints (see materials and methods). Coordinates from Met1 to Gln29 (chain A), PMP and prephenate were extracted from the MD trajectory every ps (20000 structures) and grouped into 16 clusters corresponding to coordinate root mean square deviation (rmsd) larger than 3.5Å. From the

evolution of cluster population versus time (shown in Figure 9), 6 most populated clusters, visited many times all along the simulation, seem particularly interesting and have been studied further.

Final production dynamics from selected clusters

The coordinates of the lowest energy structure in the 6 selected clusters were computed and one molecular dynamics simulation was started from each resulting minimized structure (see materials and methods). The evolution of dK-Prf is plotted in Figure 10 for the six selected clusters. One particular cluster (number 6) simulated during 85 ns, shows very large fluctuations of dK-Prf from 2.6 to more than 21.6 Å, demonstrating the high mobility of the N-terminal part of the protein. The corresponding two structures are shown in Figure 11 after superimposition of backbone atoms of residues 33 to 382. Backbone atom coordinate rmsd is 4.18 Å for residues 13 to 32 and 1.09 Å for residues 33 to 382. Between the two structures, conformational changes around Lys12 appear with transitions of the backbone dihedral angle Φ=(Met11)C-N-Cα-Cβ from 180° to 120° and the side chain dihedral angle χ1=N-Cα-Cβ-Cγ from 60° to -60°, for "far" and "close" conformation, respectively.

The presence of Lys/Arg/Gln12 could be considered as a molecular signature of the PAT competent 1β AAT

Our experimental kinetics data, revealed that Lys12 plays a specific role in the acquisition of PAT activity by *R. meliloti* and *T. thermophilus* 1β AAT most certainly by stabilizing prephenate though an interaction with its 4-hydroxyl group. Accordingly, we were able to optimize at the QM level a structure of the active site respecting this restraint. Releasing this restraint during MD simulations results in a rapid drift of dK-Prf followed by frequent returns to dK-Prf lower than 3 Å. This means that, although no crystal structure of PAT competent 1β AAT from *T. thermophilus*, or homologues, has been obtained showing proximity between Lys12 and the catalytic site, interactions are possible and even frequent between Lys12 and prephenate, on the short time scale accessible to MD simulations.

These simulations strongly support, at the structural level, our kinetics data. All together our data point lysine12 and the corresponding arginine or glutamine, present in other PAT competent 1β AAT, as key residues in the acquisition of a physiologically relevant prephenate aminotransferase activity by 1β AAT during evolution. As already mentioned, our data also revealed that other residues that remain to be identified are responsible for the good reaction rate exhibited by all tested plant and some bacteria 1β AAT PAT.

Phylogenetic distribution of 1β AAT among different classes of organisms (Figure 12) confirms the presence of a 1β AAT with a Lys/Arg/Gln in position 12 in all photosynthetic organisms at the exception of most Cyanobacteria, which are known to synthesize arogenate via a BCAT-PAT [14]. The relative distribution of K/R/Q12 reveals the presence of a Lysine residue (corresponding to K12) in the majority of land plants except in the Poaceae (like *Zea maïs* or *Sorghum bicolor*) which possess an arginine. Most green algae also have a Lysine residue, but in the 1β AAT from *Ostreococcus lucimarinus* (A4S857) the lysine is replaced by an arginine, like in Poaceae. It is difficult to have a clear idea of the situation for the brown and the red algae due to the small number of sequenced genomes. Some brown algae like *Sargassum thunbergii* have a 1β AAT with a serine residue (AOA097IUX3) and presumably synthesize arogenate with a BCAT-PAT like Cyanobacteria or a N-succinyl-l,l-diaminopimelate aminotransferase like the two tested actinobacteria, *Streptomyces avermitilis* and *Mycobacterium tuberculosis* [14], or do not synthesize tyrosine and phenylalanine via the arogenate route. In the other hand, some brown algae like *Punctaria latifolia* possess a 1β AAT (AOA097IUR7) with an arginine like Poaceae. The few sequenced genomes of red algae seem indicate that they are devoid of 1β AAT, the first hit concern aspartate aminotransferases with a highly divergent N-terminal sequence but with a conserved K101 and a more conserved C-terminal sequence. In contrast, Diatoms which result from secondary endosymbiosis events between an algae and a eukaryote, have either a Lysine like green algae, or a Q like some photosynthetic bacteria, confirming the possible dual origin (green or red algae) of their chloroplasts [19]. Intriguingly, *Nanochloropsis* have a 1β AAT with a Glu in the corresponding 12 position. Chlorobi and Bacteroidetes have a Q, and interestingly the 1β AAT of the symbiotic cyanobacteria *Synechococcus spongiarum* also have à Q and branch with Chlorobi and Bacteroidetes rather than with other Cyanobacteria. Lys12 residues are found in 1β AAT from many non-photosythetic bacteria (most α-proteobacteria and hadobacteria and some γ-proteobacteria) known to be arogenate competent organisms. In the other hand, most Actinobacteria like *Streptomyces*

avermitilis have 1β AAT with a Serine12, which is in agreement with their capacity to synthesize arogenate through an S-DAP-PAT [14].

Experimental procedures

Bacterials strains- Two strains of *E. coli* were used in this work: the strain DH5α (Invitrogen strain; F-Φ80lacZΔM15 Δ (lacZYA-argF) U169 deoR recA1 endA1 hsdR17 (rk-, mk +) phoA supE44 thi-1 gyrA96 relA1 λ-) was used for plasmid DNA amplification and subcloning. The strain BL21 Rosetta 2 (DE3) (Merck; F-ompThsdSB (rB-mB-) gal dcm (DE3) pRARE2 (CamR)) was used for overproduction of recombinant proteins. These two strains were grown at 37°C in Luria-Bertani medium supplemented with 1% agar (w/v) for the agar media, and supplemented with Carbenicillin 1‰ (DH5α) or Kanamycine 1‰ (BL21). The 2011 RCR strain of *Rhizobium meliloti* was provided by G. Alloing (UMR6192, Sophia Antipolis). *Thermus thermophilus* (strain HB8) Genomic DNA was provided by A. Atteia (UMR 7281, Marseille). *Synechococcus* CC9605 genomic DNA was provided by F. Partensky (UMR7144, Roscoff)

Chemicals- Prephenate, oxaloacetate, and glutamate were purchased from Sigma.

Primers- Primers used for amplification of the 1β aspartate aminotransferase genes from the corresponding genomic DNA are presented in Table 4. Restriction sites introduced by PCR are underlined. These cDNAs were cloned into pET30 b(+) vector, and the recombinant proteins were produced without any tag.

Overproduction and partial purification of recombinant proteins- Fresh colonies of transformed BL21 (DE3) Rosetta2 bacteria (Novagen, Darmstadt, Germany) containing pET30 b(+) vector expressing the desired recombinant protein were transferred into 15 ml of LB medium supplemented with Kanamycine 1‰, and grown at 37°C. Saturated culture was transferred into 800 ml LB medium supplemented with Kanamycine 1‰,, and growth was continued until $DO_{600\ nm}$=0.6. IPTG (0.4 mM) was added to the medium, and the bacteria were grown at 20°C for 16 h. Cells were harvested by centrifugation (4000 g, 45 min), and pellets were re-suspended in 30 ml of 25 mM HEPES pH 8.0, 1 mM EDTA, 1 mM DTT, 10% glycerol (v/v), 50 µM PLP, 5 mM ε-aminocaproic acid, and 1 mM benzamidine (Buffer A), and sonicated for 10 min at 4°C on a Branson sonicator. Streptomycin sulphate (0.1% (w/v)) was added to precipitate DNA and the extract was centrifuged for 45 min at 40 000 g (Sorvall SS-34) at 4°C. The resulting supernatant of soluble proteins were stored at -80°C until used for purification.

Partial purification of recombinant proteins from *R. meliloti* Q02635 WT and K12G mutant were performed as follows: crude extract of soluble proteins (20-50 mg) were loaded on a 30 ml of Q-sepharose resin (GE Healthcare) in an XK-16 column (Amersham Pharmacia) equilibrated with Buffer A, proteins were eluted by a linear gradient (0-0.5 M) of NaCl in buffer A. Active fractions were pooled and loaded on HiPrep 16/60 Sephacryl S-200 column (Amersham Pharmacia) equilibrated with Buffer A supplemented with 100 mM NaCl (Figure 13). The most active fractions were pooled, desalted in buffer A supplemented with 10% glycerol, frozen in liquid nitrogen and stored at -80°C.

Partial purification of recombinant proteins from *Thermus thermophilus* Q56232 WT and K12G mutant were performed as follows: crude extract of soluble proteins (20-50 mg) were loaded on 100 ml of EMD DEAE 650(M) resin Merck) in an XK 26 column (Amersham Pharmacia) equilibrated with Buffer A, proteins were eluted by a linear gradient (0-0.5 M) of NaCl in buffer A (Figure 14). The most active fractions were pooled, desalted in buffer A supplemented with 10% glycerol, frozen in liquid nitrogen and stored at -80°C.

Partial purification of recombinant proteins from *Synechoccocus* Q3AK33 were performed as follows: crude extract of soluble proteins (20-50 mg) were loaded on a 30 ml of Q-sepharose resin (GE Healthcare) in an XK-16 column (Amersham Pharmacia) equilibrated with Buffer A, proteins were eluted by a linear gradient (0-0.5 M) of NaCl in buffer A (Figure 5). The most active fractions were pooled, desalted in buffer A supplemented with 10% glycerol, frozen in liquid nitrogen and stored at -80°C

Determination of enzyme activities- Glutamate-oxaloacetate aminotransferase activities were assayed as described in [8].

Prephenate aminotransferase activities were assayed by coupling the reaction with purified Tyr-insensitive arogenate-specific dehydrogenase from *Synechocystis* [22] and following the reduction of NADP at 340 nm. The reaction was carried out at 30°C in 50 mM HEPES buffer (pH 8.0) in the presence of 40 nM coupling enzyme, 100 µM NADP and variable amounts of prephenate, and fixed concentration of glutamate (25mM). Activities were calculated using an epsilon for NADPH of 6250 $M^{-1}.cm^{-1}$ at 340 nm. Arogenate formation was confirmed by HPLC analyses as described in [10].

Kinetic data analysis- Initial velocity data were fitted to the Michaelis–Menten equation using the Kaleidagraph program (Synergy Software, Reading, PA, USA).

Crystallization- Extensive searches for crystallization conditions were performed in 96 well plates at 293 K using the high throughput crystallization facilities of EMBL (Grenoble). Crystallization hits were refined in 24 well Linbro plates. All the crystals were grown at 293 K. Ath-PAT crystals grew in 0.1 M Na citrate pH 4.0, 11 % PEG 4K. Rme-AAT/PAT crystals were obtained in 0.2 M ammonium formate, 19 % PEG 3350 and Rme-AAT grew in 0.1 M Na acetate pH 4.5, 5% PEG 4K. The protein concentration was 10 mg/ml.

Data collection- Diffraction data (Table 1) were collected at 100 K at the European Synchrotron Facilities either on FIP-BM30A [23] (Rme-AAT) or ID23-1 (Ath-PAT and Rme-AAT/PAT) using an ADSC 315r detector. Prior freezing, the crystals were soaked in mother liquor with glycerol in a concentration range from 20 % (Ath-PAT and Rme-AAT/PAT) to 25 % (Rme-AAT). All the data were processed and scaled using XDS [24].

Phasing and Refinement- All the phase calculations were performed by molecular replacement after modifying the models using CHAINSAW [25] from CCP4 [26]. Phases for Ath-PAT were calculated using Molrep [27] and the atomic coordinates of aspartate aminotransferase from *Phormidium lapideum* (PDB code 1J32) as a model. Phaser [28] was used to calculate phases for Rme-AAT/PAT and Rme-AAT using the atomic coordinates of Ath-PAT and those of the aspartate aminotransferase from *Phormidium lapideum* (PDB code 1J32), respectively. All the refinements and rebuilding were carried out using Phenix [29] and COOT [30]. Non-crystallographic symmetry was applied in all cases during refinement. Rmsd calculations were done using PDBefold [31].

Atomic coordinates and X-ray data were deposited in the PDB with the accession numbers: At-PAT (6F5V), Rme-PAT (6F77) and Rme-AspT (6F35).

Structural modeling

I- Modelling of Lys12 conformational transitions using MD simulations

Initial model- It was built from the structure of aspartate aminotransferase from *Thermus thermophilus* (Tth-PAT) complexed with maleate (PDB id 1BKG). 1BKG gives the structure from Met1 to Leu382 in both A and B chains. It was obtained in the presence of 4'-deoxy-4'-aminopyridoxal-5'-phosphate (PMP) and maleate.

Experimentalists have modified the structure "manually" changing the position of Lys12 "by eye" to fit experimental observations and introducing global rotations and translations with respect to 1BKG to obtain a structure called Sexp. They also manually inserted an ideal structure of prephenate in place of maleate. The global atomic coordinate rms deviation between Sexp and 1BKG is 0.85 Å (considering the 5888 atoms of the dimer without water and ligands). To start from the experimental structure 1BKG without losing the information coming from experiments, we started by calculating the reverse transformation to be applied to Sexp to bring it back close to 1BKG while maintaining the rearrangement of Lys12. Then, this transformation was applied to prephenate and crystal waters, as well.

Structure preparation- Finally the structure of the dimer of Tth-PAT from residue 1 to 382, PMP, prephenate and 141 crystal waters was completed and refined with the molecular dynamics program CHARMM [32]. The all-atom force field *all27* for proteins [33,34] was used. An initial structure of prephenate was obtained from the Zinc database (zinc1532584) while the initial structure of pyridoxamine

phosphate (PMP) was loaded into chimera with its PubChem ID (cid1053). Both structures were saved to mol2 format and submitted to the SwissParam server which provided initial topology and parameter files for CHARMM. All histidines were protonated on their Nδ atom, whereas aspartate and glutamate residues were kept unprotonated (negative charge). All arginine and lysine residues were positively charged. Missing hydrogen atom coordinates were built with CHARMM and the structure was energy minimized down to a gradient of 0.1 kcal/mol/Å with the ABNR algorithm and subject to harmonic restraints on heavy atoms with force constant of 5 kcal/mol/Å^2. This led to an initial refined structure with a dK-Prf of 2.9 Å. To complete the initial unrestrained model, a 1 ns Langevin Molecular Dynamics simulation at a temperature of 300 K was run on this system containing the 1BKG dimer, the X-ray waters, prephenate and PMP. (The frame giving the highest CHARMM interaction energy between prephenate or PMP and the protein is shown in Figure 3).

Solvation and restrained optimization- Twelve sodium ions were placed around the protein at initial locations minimizing their interaction energy with the protein to neutralize the system. Then the system was solvated inside a 109*66*66 Å^3 water box initially containing 15886 waters molecules. Periodic boundary conditions were applied in the orthorhombic symmetry. The protein atoms were initially fixed, the crystal water oxygen atoms were harmonically restrained with a force constant of 2.0 kcal/mol/Å^2 and the solvent around the protein was equilibrated with 5000 steps (10 ps) Langevin Molecular Dynamics simulation at a temperature of 300 K yielding structure *1BKG_Preph_Mod* (after introduction of prephenate into 1BKG, solvation and energy minimization).

Then all protein atoms were set free to move, and harmonic restraints were introduced on protein backbone atoms, crystal water oxygens and heavy atoms of PMP with a force constant of 1.0 kcal/mol/Å^2. Additionally, Nuclear Overhauser Enhancement (NOE)-type distance restraints using a bi-harmonic potential with a force constant of 40 kcal/mol/Å^2 were applied between:

- Lys12 NZ atom and prephenate O1 oxygen, dK-Prf (above 2.0 and below 3.5 Å)
- PMP H1 or H2 atoms and prephenate O2 oxygen (above 1.0 and below 3.5 Å)
- Lys101 NZ atom and prephenate O6 oxygen (above 2.0 and below 3.5 Å)
- Arg361 HH11 and HH21 atoms and prephenate O3 or O5 oxygen (above 1.0 and below 3.0 Å)

The system was energy minimized and subject to 10 more ps Langevin MD.

Equilibration- Starting with the equilibration phase, electrostatics was calculated using the Particle Mesh Ewald summation and the protein center of mass was restrained to the center of the coordinate system. The NOE-type distance restraints previously described were applied together with harmonic restraints on protein backbone atoms and heavy atoms of PMP with a force constant of 1.0 kcal/mol/Å^2. Two rounds of steepest descent followed by ABNR minimization were run, the latter acting on the box dimensions as well as the atom coordinates. The system was then equilibrated for 100 ps using Langevin dynamics at 300 K. Harmonic restraints on the protein and PMP were removed leaving NOE-restraints only. Then, the system was subject to a constant pressure of 1 atm and temperature of 300 K and a constant pressure and temperature CPT (Constant Pressure and Temperature) dynamics based on Berendsen's weak coupling method [35] was run for 20 ns leading to Structure *1BKG_Preph_MD* (after introduction of prephenate into 1BKG, solvation, equilibration and 20 ns production Molecular Dynamics simulation).

II-Parameter optimization

Partial charge optimization through QM/MM study- A refined parameterization of the two substrates: prephenate and PMP and surrounding residues was done. Indeed, the atomic charges given by SwissParam are probably deduced from small compounds and certainly do not take into account the specific environment of the molecules. Starting from a structure with low energy and low dK-Prf (3.2Å) from previous equilibration, we have run a QM/MM optimization followed by a calculation of the charges adequate to fit the calculated electrostatic potential around the ligands.

We have used the Gaussian03 package [36] for the quantal parts (QM parts hereafter) and Tinker 4.2 [37] for the part described by molecular mechanics (MM part). In addition to prephenate PMP and residue Lys12, the QM part consisted of side chains of Lys101, Trp125, Ser127, Lys234, Arg242, Tyr322, Arg361 from chain A and Tyr64 from chain B. 14 water molecules were also included in this quantum part for a total of 191 QM atoms. All MM atoms belonging to residues within a distance of 3 Å of the QM part were allowed to relax during geometry optimizations, and the remaining structure was kept frozen. Dangling

bonds between the QM and MM subsystems were treated using the Link-Atom scheme [38] as implemented in Ref. [39].

For the QM part, the B3LYP density functional method [40,41,42] with a 6-31G* basis set for all QM atoms was used, which represents 1484 basis functions. The MM part was treated with the ChARMM27 force field. Missing residues and hydrogens were added to complete the structure. From this energy-minimized structure, charges on atoms in the active region were calculated using the Merz-Kollman routine [20] in Gaussian to fit the *ab initio* electrostatic potential. After checking that the calculated charges for water atoms were very close to those used in the CHARMM MM force field, these charges were directly transferred to the protein structure file (psf) describing the MM system for prephenate, PMP and surrounding residues.

New molecular dynamics simulations- Using this modified structure file, 24 ns Constant Pressure and Temperature dynamics were run restarting from structure *1BKG_Preph_MD*. The NOE (Nuclear Overhauser Enhancement) type restraints used to obtain *1BKG_Preph_MD* were kept for four ns and then removed for the remaining 20 ns.

Improving conformational sampling- A method to explore more conformational space is to use Replica eXchanging Self-Guided Langevin Dynamics as implemented in CHARMM in the RXSGLD module [21]. SGLD simulations enhance conformational search efficiency through acceleration of low frequency motions in a molecular system. The low frequency motion represents the slow conformational change that contributes the most in conformational search. A guiding force resonant with the low frequency motion, which is calculated based on a local average of momentums, is introduced into the equation of motion to enhance the low frequency motion. Here, temperatures of replicas and their self-guiding temperatures were exponentially spaced from 300 to 400 K and from 300 to 600 K, respectively. The guiding factor increment was set to 0.2 and exchange was attempted every 50 steps [43]. 16 replicas were created starting from structure *1BKG_Preph_Mod* and vacuum RXGLD simulations were run for 10 ns with a time step of 1 fs on 16 processors. To explore conformational transitions compatible with a proximity between Lys12 and prephenate, NOE-type distance restraints on Lys12, prephenate and PMP were turned on and protein backbone atom coordinates were harmonically restrained with force constant of 2.0 kcal/mol/Å^2 except for the 29 N-Terminal amino acids of monomer A which were left completely free to move. Coordinates of Met1 to Gln29, chain A and PMP + prephenate were extracted from the MD trajectory every ps (20000 structures) and grouped into 16 clusters corresponding to coordinate rmsd larger than 3.5 Å.

Final production dynamics from selected clusters- The coordinates of the lowest energy structure in the 6 selected clusters were computed and one molecular dynamics simulation was started from each resulting minimized structure. Each system was neutralized, solvated and equilibrated for 100 ps with 4 NOE-type distance restraints as explained above. Harmonic restraints on the protein and PMP were removed leaving NOE-restraints only. Then, for the production phase, the system was subject to a constant pressure of 1 atm and temperature of 300 K and a CPT dynamics based on Berendsen's weak coupling method [35] was run for 5 ns. Then the NOE-restraint on Lys12 was removed and the dynamics was continued.

Acknowledgments

The diffraction experiments were conducted on beamline FIP-BM30A and ID23-1 at the ESRF (Grenoble, France). We thank the beamline staff for technical help. This work used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

This work was supported by the French National Institute for Agricultural Research (INRA) (post doc fellowship to AR), the Centre National de la Recherche Scientifique (CNRS), the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), and the University of Grenoble Alpes (PhD fellowship to MG, and a grant from the Grenoble Alliance for Integrated Structural Cell Biology: ANR-10-LABEX-04).

Footnote

*During the final stage of writing our Manuscript, a new structural study on *Arabidopsis* prephenate amino transferase appeared online in plant journal (14). This study confirms that Ath-PAT behaves very similarly to Tth-PAT during substrate binding.

References

1- Haslam, E (1993) *Shikimic acid, metabolism and metabolites*, Chichester, UK

2- Razal, RA, Ellis, S, Singh, S, Lewis, N, & Towers, GHN (1996) Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry* 41, 31-35

3- Dewick, PM (1998) The biosynthesis of shikimate metabolites. *Nat Prod Rep* 15, 17-58

4- Stenmark, SL, Pierson, DL, & Jensen, RA (1974) Blue-green bacteria synthesize L-tyrosine by pretyrosine pathway. *Nature* 247, 290-292

5- Fazel, AM, & Jensen, RA (1979) Obligatory biosynthesis of L-tyrosine via the pretyrosine branchlet in coryneform bacteria. *J Bacteriol* 138, 805-815

6- Jung E, Zamir LO, & Jensen RA (1986) Chloroplasts of higher plants synthesize L-phenylalanine via arogenate. *Proc Natl Acad Sci USA* 83: 7231-7235.

7- Rippert, P., & Matringe, M. (2002) Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of *Arabidopsis thaliana*. *Eur. J. Biochem.* 269, 4753–4761

8- Cho M-H, Corea ORA, Yang H, Bedgar DL, Laskar DD, Anterola AM, Moog-Anterola FA, Hood RL, Kohalmi SE, Bernards MA, Kang CH, Davin LB, & Lewis NG (2007) Phenylalanine Biosynthesis in Arabidopsis thaliana Identification and characterization of arogenate dehydratases. *J. Biol. Chem*., 282, 30827-30835

9- Christendat, D, & Turnbull, JL (1999) Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from Escherichia coli. *Biochemistry* 38, 4782-4793

10- Graindorge, M, Giustini, C, Jacomin, AC, Kraut, A, Curien, G, & Matringe, M (2010) Identification of a plant gene encoding glutamate/aspartate-prephenate aminotransferase: The last homeless enzyme of aromatic amino acids biosynthesis. *FEBS Lett* 584, 4357-4360

11- Maeda, H, Yoo, HJ, & Dudareva, N (2011) Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nat ChemBiol* 7, 19-21

12- Jensen, RA, & Gu, W (1996) Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. *J Bacteriol* 178, 2161-2171

13- Abouzeid, A, Euverink, GJW, Hessels, GI, Jensen, RA, & Dijkhuizen, L (1995) Biosynthesis of L-phenylalanine and L-tyrosine in the actinomycete *Amycolatopsis methanolica*. *Applied Env Microbiol* 61, 1298-1302

14- Graindorge, M, Giustini, C, Kraut, A, Moyet, L, Curien, G, & Matringe, M (2014) Three different classes of aminotransferases evolvedprephenate aminotransferase functionality in arogenate-competent microorganisms *J Biol Chem*., 289, 3198-31208

15- Nakai, T, Okada, K, Akutsu, S, Miyahara, I, Kawaguchi, S, Kato, R, Kuramitsu, S, and Hirotsu, K (1999) Structure of *Thermus thermophilus* HB8 aspartate aminotransferase and its complex with maleate. *Biochemistry*, 38, 2413-2424

16- Holland, CK, Berkovich DA, Kohn, ML, Maeda, H, & Jez, JM (2018) Structural basis for substrate recognition and inhibition of prephenate aminotransferase from Arabidopsis. *Plant Journal* Apr;94(2):304-314

17- Dornfeld, C, Weisberg, AJ, Ritesh, KC., Dudareva, N, Jelesko, JG., & Maeda, HA. (2014). Phylobiochemical characterization of class-Iβ aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *Plant Cell*, 3101-3114

18- Siehl, DL., Connelly, JA. & Conn, EE (1986). Tyrosine biosynthesis in *Sorghum bicolor*: characteristics of prephenate aminotransferase. Z Naturforsch C. ;418(1):79–86.

19- Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, & Bhattacharya D. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 26; 324(5935):1724-1726.

20- Besler, BH., Merz, KM Jr & Kollman, PA (1990) Atomic charges derived from semi-empirical methods. *J. Comput. Chem*., 11, 431-439
21- Wu, X, Hodoscek, M, & Brooks, BR (2012) Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling. *J. Chem. Phys.* 137044106
22- Legrand, P, Dumas, R, Seux, M, Rippert, P, Ravelli, R, Ferrer, J L, & Matringe, M (2006) Biochemical characterization and crystal structure of *Synechocystis* arogenate dehydrogenase provide insights into catalytic reaction. *Structure*, 14, 767-776
23- Roth, M, Carpentier, P, Kaïkati, O, Joly, J, Charrault, P, Pirocchi, M, Kahn, R, Fanchon, E, Jacquamet, L, Borel, F, Bertoni, A, Israel-Gouy, P & Ferrer, JL (2002). FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Cryst*., D 58, 805-814
24- Kabsch, W (2010). XDS. *Acta Cryst*., D 66, 125-132
25- Stein, N, (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst*., 41, 641-643
26- CCP4. Collaborative computational project, Number 4. The CCP4 suite: programs for protein crystallography. *ActaCryst.*, D 50, 760-763
27- Vagin, A, & Teplyakov A (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, 30, 1022-1025
28- McCoy, AJ, Grosse-Kunstleve, RW, Adams, PD, Winn, MD, Storoni, LC & Read, RJ (2007). *J. Appl. Cryst*., 40, 658-674
29- Adams, PD, Afonine, PV, Bunkóczi, G, Chen, VB, Davis, IW, Echols, N, Headd, JJ, Hung, LW, Kapral, GJ, Grosse-Kunstleve, RW, McCoy, AJ, Moriarty, NW, Oeffner, R, Read, RJ, Richardson, DC, Richardson, JS, Terwilliger, TC. & Zwart, PH (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *ActaCryst.*, D 66, 213-221
30- Emsley, P, Lohkamp, B, Scott WG. & Cowtan, K (2010). Features and development of Coot. *ActaCryst.*, D 66, 486-501
31- Krissinel, E & Henrick, K (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *ActaCryst.*, D 60, 2256-2268
32- Brooks, BR, Bruccoleri, RE, Olafson, BD, States, DJ, Swaminathan, S, & Karplus, M (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations.*J. Comput. Chem.*, 4, 187-217
33- MacKerell, AD, Bashford, D, Bellot, M, Dunbrack, R, Evanseck, J, Field, M, Fischer, S, Gao, J, Guo, H, Ha, S, Joseph-McCarthy, D, Kuchnir, L, Kuczera, K, Lau, F, Mattos, C, Michnick, S, Ngo, T, Nguyen, D, Prodhom, B, Reiher III, W, Roux, B, Schlenkrich, M, Smith, J, Stote, R, Straub, J, Watanabe, M, Wiorkiewicz-Kuczera, J, & Karplus, M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chems.* B 102, 3586-3616
34- Mackerell, AD, Feig, M, and Brooks, CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25, 1400-1415
35- Berendsen, H, Postma, J, van Gunsteren, W, DiNola, A, & Haak, J (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684-3690
36- Frisch, MJ, Trucks, GW, Schlegel, HB, Scuseria, GE, Robb, MA, Cheeseman, JR, Zakrzewski, VG, Jr., JAM, Stratmann, RE., Burant, JC., Dapprich, S, Millam, JM, Daniels, AD, Kudin, KN, Strain, MC, Farkas, O, Tomasi, J, Barone, V, Cossi, M, Cammi, R, Mennucci, B, Pomelli, C, Adamo, C, Clifford, S, Ochterski, J, Petersson, GA, Ayala, P., Cui, Q, Morokuma, K, Rega, N, Salvador, P, Dannenberg, JJ, Malick, DK, Rabuck, AD, Raghavachari, K, Foresman, JB, Cioslowski, J, Ortiz, JV, Baboul, AG, Stefanov, BB, Liu, G, Liashenko, A, Piskorz, P, Komaromi, I, Gomperts, R, Martin, RL, Fox, DJ, Keith, T, Al-Laham, MA, Peng, CY, Nanayakkara, A, Challacombe, M, Gill, PMW, Johnson, B, Chen, W, Wong, MW, Andres, JL, Gonzalez, C, Head-Gordon, M, Replogle, ES, & Pople, JA Gaussian 98, Revision A.11.3. Gaussian, Inc., Pittsburgh PA, 2002
37- Ponder, JW (2001). TINKER: Software Tools for Molecular Design, Version 4.2.
38- Field, M, Bash, P, & Karplus, MA (1990). Combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 11, 700-733
39- Ferre, N & Olivucci, M (2003) Probing the rhodopsin cavity with reduced retinal models at the CASPT2//CASSCF/AMBER level of theory. *J. Am. Chem. Soc.* 125, 6868-6869

40- Becke, AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098-3100
41- Lee, C, Yang, W, & Parr, RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter*, 37,785- 789
42- Becke, AD (1993) Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J.Chem. Phys.* 98, 5648-5652
43- Jiang, W, Hodoscek, M, & Roux, B (2009) Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics (FEP/REMD). *J. Chem. Theory. Comput.* 5, 2583-2588

Author contribution

All the authors have contributed to the writing of the manuscript. CG and MG performed biochemical and crystallization experiments, DC and AR solved the 3D structures, SG performed molecular modeling, DC, SC, GC and MM designed the research

Table 1: data collection statistics

	Ath-PAT	Rme-PAT	Rme-AAT
Wavelength (Å)	0.97967 (FIP/BM30A)	0.97931(ID23-1)	0.97970 (FIP/BM30A)
Resolution (Å)	23.70-1.70 (1.74-1.70)	47.30-1.79 (1.84-1.79)	47.75-1.90 (1.95-1.90)
Unit Cell (Å; °)	58.47 73.24 103.32 α=92.49 β=87.22 γ=111.05	103.54 93.01 123.07 β=91.36	72.20 117.28 127.30
Space group	P1	$P2_1$	$P2_12_12$
Molecules in au	4	6	2
Total reflections	1,035,603 (42,878)	632,614 (97,520)	551,325 (38,783)
Unique reflections	168,766 (11,501)	213,629 (33,152)	85,690 (6,128)
Redundancy	6.14 (3.73)	2.96 (2.94)	6.43 (6.33)
Completeness (%)	96.00 (88.20)	98.0 (94.40)	99.80 (98.10)
Rsym (%)	6.40 (48.40)	9.40 (57.70)	10.70 (60.90)
Rmeas (%)	6.9 (56.6)	11.7 (71.0)	11.7 (66.5)
I/σ_I	16.57 (2.75)	9.77 (2.28)	14.64 (3.27)

CC1/2	99.9 (82.9)	99.5 (72.9)	99.8 (86.9)

Rsym=$\frac{\sum\sum|I_i-I_m|}{\sum\sum I_i}$, where I_i is the intensity of the measured reflection and I_m is the mean intensity of this reflection. Values indicated in parentheses correspond to the statistics in the highest resolution shell.

Table 2: Refinement statistics

	Ath-PAT	Rme-PAT	Rme-AAT
Resolution (Å)	23.70-1.70 (1.72-1.70)	47.30- 1.79 (1.81-1.79)	47.75-1.90 (1.92-1.90)
R_{cryst} (σ_F=0) (%)	17.23 (26.58)	17.44 (29.39)	15.25 (21.96)
R_{free} (σ_F=0) (%)	20.12 (25.87)	20.87 (33.67)	18.07 (25.83)
Number of atoms	13,825	19,872	7,127
Water molecules	1,232	1,616	851
B average (Å^2)	31.0	29.0	21.0
Rmsd bonds (Å)	0.006	0.007	0.006
Rmsd angle (°)	1.075	0.876	0.859
Ramachandran statistics			
Residues in allowed zones (%)	99.8	99.8	100.0
Residues in disallowed zones (%)	0.0	0.0	0.0

Values indicated in parentheses correspond to the statistics in the highest resolution shell.

Table 3: kinetics analyses of 1b AAT Wild Type and K12G mutants from *R. meliloti* (Q02635) and *T. thermophilus* (Q56232).

Rme-PAT	$K_M^{prephenate}$(µM)	$k_{cat\ Prephenate}$(s^{-1})	$K_{M\ Oxaloacetate}$ (µM)	$k_{cat\ Oxaloacetate}$ (s^{-1})
WT	106±12	43±8.2	13.7 ±2.4	168±22.3
K12G	692±143	39.5±6.2	21.8 ±3.4	156±23.6
Tth-PAT	$K_{M\ prephenate}$(µM)	$k_{cat\ Prephenate}$(s^{-1})	$K_{M\ Oxaloacetate}$(µM)	$k_{cat\ Oxaloacetate}$(s^{-1})
WT	150±21	7.7±1.3	25.3±3.1	32.2±4.3

K12G	2323±250*	5.2±2.5*	26.4±2.8	12.3±1.7

Values given are the average of at least three independent determinations from a single enzyme preparation ± SD. *Note that K12G mutant kinetic apparent parameters K_M and k_{cat} for prephenate were difficult to determine accurately due to the important deterioration of the apparent K_M.

Table 4: primers used for amplification of *Rhizobium meliloti* (Q02635), *Thermus thermophilus* (Q56232) *and Synechochoccus* Q3AK33 1βAAT, and primers used for directed mutagenesis of K12G mutants.

Enzyme	Primer 5’
Rme-PAT	CAGGAAACACCATATGGCCTTCCTTGC
Rme-PAT K12G	CCCGTGTAGGGCCCTCCGCCACC
Tth-PAT	CCCGCGTATACTTAGCATATGCGCGGCCTTT
Tth-PAT K12G	GGTCCTAGGCCATGGGGCCCTCGGCCACG
Sco-PAT	TTGCATATGCCGCGCCCGCCAGAACTTTCAGACCGAGCTG
Enzyme	Primer 3’
Rme-PAT	CATGTCGGTCTGCTCGAGAATTATCTGC
Rme-PAT(GPSAT)	GGTGGCGGAGGGCCCTACACGGG
Tth-PAT	CCCCAGGAAAACGTCGACTAGGCGCGCCCCA
Tth-PAT(GPSAT)	CGTGGCCGAGGGCCCCATGGCCTGGACC
Sco-PAT	GAAGTGATGAGCTCGGATCAAAAGGAACGAAGCAGCCGCTTC

Figures Legends

Figure 1. Tyrosine biosynthesis pathway. The alternative routes for the synthesis of tyrosine are represented. Synthesis of phenylalanine from arogenate occurring in plants is also represented. Multiple enzymatic steps involved are shown by dotted lines. ADH: arogenate dehydrogenase; ADT: arogenate dehydratase; E4P: erythrose 4-phosphate; PAT: prephenate aminotransferase, PDH: prephenate dehydrogenase; PEP: phosphoenolpyruvate; TYR-AT: tyrosine aminotransferase

Figure 2: Three dimensional structure comparison of PAT competent and incompetent 1βAAT.
A- 3D structure of *A. thaliana* PAT competent 1βAAT dimer with PLP in yellow sticks. B- Superimposition of 3D structures of a monomer of *A. thaliana* PAT competent 1βAAT (green), and *R. meliloti* PAT competent 1βAAT (blue) and *R. meliloti* PAT incompetent 1βAAT (red).

Figure 3: Docking of prephenate in the active site of *Thermus thermophilus* PAT competent 1βAAT (PDB id 1BKG)
A Structure of the three ketoacid substrates of PAT competent 1β AAT, oxaloacetate (PDB ID 3PDB) (blue), α ketoglutarate (build from prephenate) (cyan), and prephenate (green).
B Initial docking of prephenate in the active site of Tth-PAT (PDB id 1BKG). Prephenate was inserted manually in the active site of 1BKG in place of maleate and the resulting structure energy minimized. Then a short 1 ns unrestrained MD simulation was run. The structure giving the highest interaction energy between prephenate or PMP and the protein is displayed. The two carboxylate groups of prephenate interact with Lys101 and Arg361, and its keto group points in direction of the NH2 of PMP and the catalytic Lys234. In this position, the hydroxyl group of prephenate faces a $_{10}$AMKPSA$_{15}$ amino-acid motif of the N-terminal flexible loop. In this motif Lys12 is only present in PAT competent enzymes. In the maleate-PMP-Enzyme ternary complex (1BKG), the side chain of Lys12 points outside the active site (10.5 Å from the prephenate OH group). In this simulation E40 and T16 were close to the active site but not in interaction with either PMP or prephenate.

Figure 4: N-terminal sequence comparison of PAT competent (PAT) and PAT incompetent (AAT) 1β AAT. *Arabidopsis thaliana* (Q42522): A. thal PAT, *Sorghum bicolor* (C5YYD7): S. bico PAT, *Thermus thermophiles* (Q56232): T. ther PAT, *Chlorobium.tepidum* (Q8KDS8): C. tepi PAT, *Rhodobacter sphaeroides* (A3PMF8): R. spha PAT, *Rhizobium meliloti* (Q02635): R. meli PAT, *Rhizobium meliloti* (P58350): R. meli AAT, *Streptomyces avermitilis* (Q82DR2): S. aver AAT and *Synechocystis* (Q55128): Synech AAT.

Figure 5: Analysis of recombinant *Synechochoccus* Q3AK33 1βAAT. A: N-terminal sequence alignment between *Synechochoccus* Q3AK33 and *Synechocystis* Q55128 1βAAT. B: SDS-PAGE of recombinante Sco_Q3AK33 Q sepharose purification: 1 total protein extract 20 µg; 2: Soluble protein extract 20 µg; 3 most active fraction eluted from a Qsepharose column 10 µg. C: Kinetic characterization of *Synechochoccus* Q3AK33 1βAAT. Glutamate 25mM was used as amino donor**.** Values given are the average of at least three independent determinations from a single enzyme preparation ± SD.

Figure 6: View of the QM active site around prephenate and PMP after QM/MM optimization with Gaussian/Tinker. QM waters, W125, S127 and MM atoms are omitted for clarity. Hydrogen bonds between protein and ligands are indicated by black dashes. Partial charges of the different functional groups are indicated (PREPH, PMP and Tyr hydroxyls, PREPH carboxylates and carbonyl, Lys ammoniums, Arg guadinidium).

Figure 7: **View of QM and QM/MM parts of the QM/MM simulation in an orientation close to Fig. 3**. The QM part (shown in licorice) includes several residues around the active site with PREPH and PMP (shown in Fig. 5) plus W125 and S127 side chains and 14 water molecules shown in balls and sticks. All other represented atoms are MM atoms belonging to residues within a distance of 3 Å of the QM part and allowed to relax during geometry optimizations, while all other atoms of the enzyme were kept frozen.

Figure8: Evolution of dK-Prf vs. time during a constant Pressure and Temperature dynamics. The simulation starts with the old set of charges and NOE restraints until 20 ns (S0). Then, charges were modified according to QM/MM parameterization and the simulation was restarted for 4 ns. Finally, all NOE restraints were removed for the remaining 20 ns (24 to 44 ns).

Figure 9: Evolution of cluster population versus time, following 16 20 ns RXSGLD simulations. Some clusters like cluster 1 or 7 appear at the beginning of the simulation and then disappear while others show up after 14 ns like cluster 15. Others like 2, 4, 8, 10, 12 or 14 appear only transiently. Clusters 1, 3, 6, 9, 13 and 15 are the most populated (>= 5%).

Figure 10: Evolution of dK-Prf vs. time during constant Pressure and Temperature dynamics simulations on selected clusters. The Noe restraints were removed after 5 ns. (Traces were smoothed by moving averages over a 20 point window). Most simulations show a rapid increase of dK-Prf just after the release of the restraints (5 ns). Although longer simulation times would be necessary to follow each of them, cluster 6 shows a decrease of dK-Prf down to less than 4 Å, already after 22 ns. During the 85 ns simulation of this cluster dK-Prf presents very large fluctuations from 2.6 to more than 21.6 Å.

Figure 11: Superimposition of the two structures with largest difference of dK-Prf (Lys12 NZ / prephenate O1 oxygen distance) in the RXSGLD simulation of cluster 6. Structure with dK-Prf= 2.6 Å is shown in cyan and structure with dK-Prf= 21.6 Å is shown in green. The conformational change of Lys12 is clearly visible (red arrow).

Figure 12: Molecular Phylogenetic analysis of 1βAAT by Maximum Likelihood method
The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-16448,6813) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 361 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Nature of the amino acid residue in position 12 (according to *Thermus thermophilus* 1βAAT) is represented.

Figure 13: SDSPAGE of the partial purification of recombinant *Rhizobium meliloti* 1βAAT PAT (Q02635) wild type and K12G mutant.
MW: molecular weight; T: total protein extract; S: soluble protein extract; Q: Q sepharose pool; S200: S200 pool.

Figure 14: SDS-PAGE of EMD DEAE partial purification of recombinant *Thermus thermophilus* 1βAAT Q56232 wild type and K12G mutant: T: total protein extract 20 µg; S: Soluble protein extract 20 µg. Arrows indicate the most active fraction used for kinetic analysis.

Figures

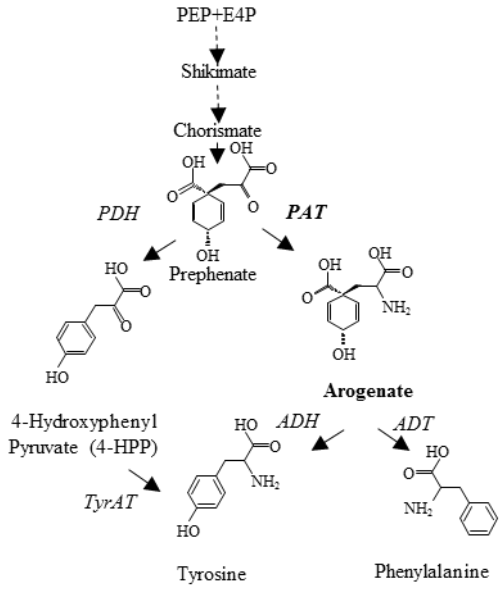


Figure 1. Tyrosine biosynthesis pathway. The alternative routes for the synthesis of tyrosine are represented. Synthesis of phenylalanine from arogenate occurring in plants is also represented. Multiple enzymatic steps involved are shown by dotted lines. ADH: arogenate dehydrogenase; ADT: arogenate dehydratase; E4P: erythrose 4-phosphate; PAT: prephenate aminotransferase, PDH: prephenate dehydrogenase; PEP: phosphoenolpyruvate; TYR-AT: tyrosine aminotransferase

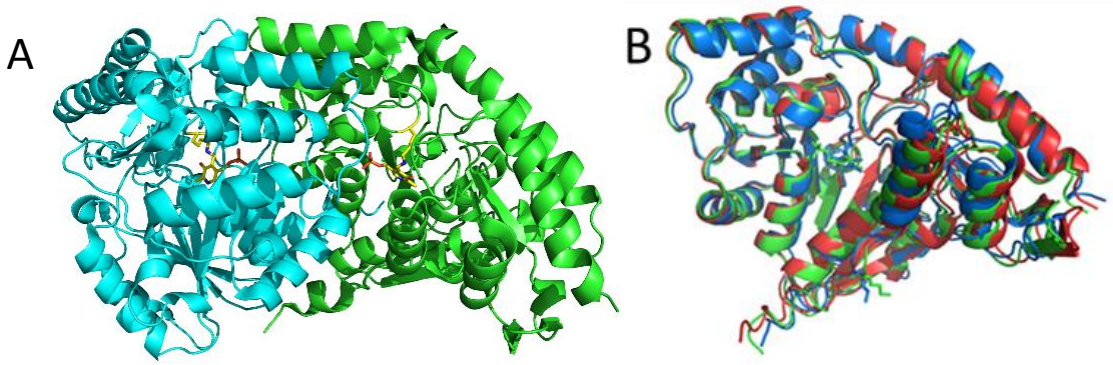


Figure 2: Three dimensional structure comparison of PAT competent and incompetent 1βAAT.
A- 3D structure of *A.thaliana* PAT competent 1βAAT dimer with PLP in stick. B- Superimposition of 3D structures of a monomer of *A. thaliana* PAT competent 1βAAT (green), and *R. meliloti* PAT competent 1βAAT (blue) and *R. meliloti* PAT incompetent 1βAAT (red).

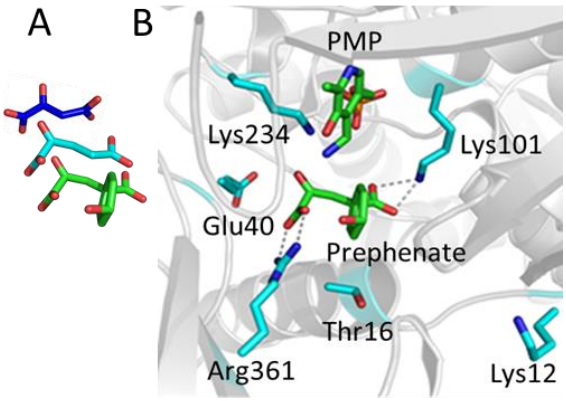


Figure 3: Docking of prephenate in the active site of *Thermus thermophilus* PAT competent 1βAAT (PDB id 1BKG)
A Structure of the three ketoacid substrates of PAT competent 1β AAT, oxaloacetate (PDB ID 3PDB) (blue), α ketoglutarate (build from prephenate) (cyan), and prephenate (green).
B Initial docking of prephenate in the active site of Tth-PAT (PDB id 1BKG). Prephenate was inserted manually in the active site of 1BKG in place of maleate and the resulting structure energy minimized. Then a short 1 ns unrestrained MD simulation was run. The structure giving the highest interaction energy between prephenate or PMP and the protein is displayed. The two carboxylate groups of prephenate interact with Lys101 and Arg361, and its keto group points in direction of the NH2 of PMP and the catalytic Lys234. In this position, the hydroxyl group of prephenate faces a $_{10}$AMKPSA$_{15}$ amino-acid motif of the N-terminal flexible loop. In this motif Lys12 is only present in PAT competent enzymes. In the maleate-PMP-Enzyme ternary complex (1BKG), the side chain of Lys12 points outside the active site (10.5 Å from the prephenate OH group). In this simulation E40 and T16 were close to the active site but not in interaction with either PMP or prephenate.

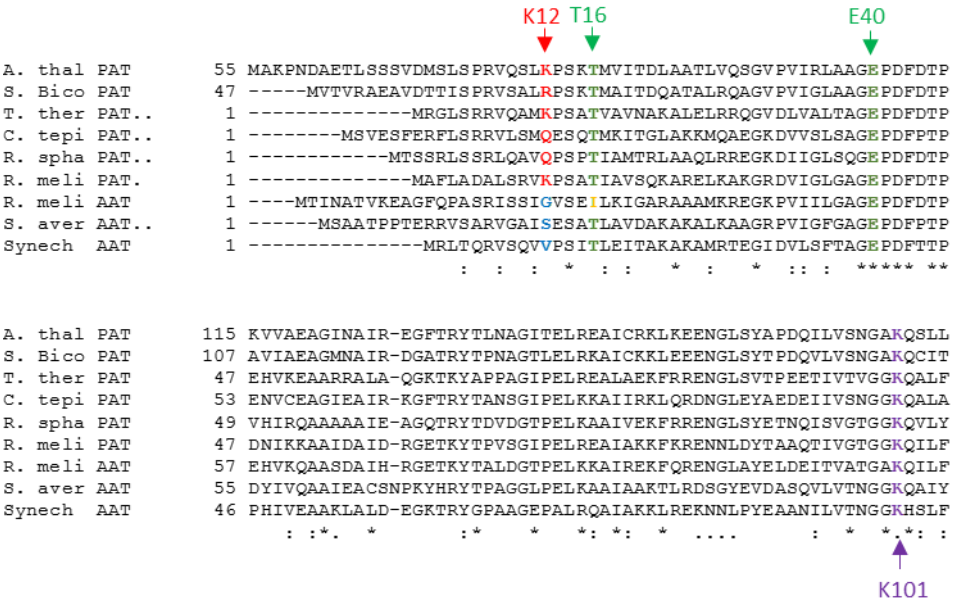
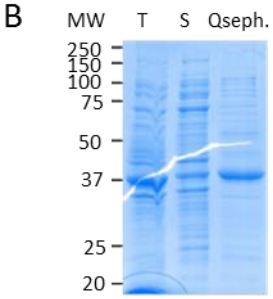


Figure 4: N-terminal sequence comparison of PAT competent (PAT) and PAT incompetent (AAT) 1β AAT. *Arabidopsis thaliana* (Q42522): A. thal PAT, *Sorghum bicolor* (C5YYD7): S. bico PAT, *Thermus thermophiles* (Q56232): T. ther PAT, *Chlorobium.tepidum* (Q8KDS8): C. tepi PAT, *Rhodobacter sphaeroides* (A3PMF8): R. spha PAT, *Rhizobium meliloti* (Q02635): R. meli PAT, *Rhizobium meliloti* (P58350): R. meli AAT, *Streptomyces avermitilis* (Q82DR2): S. aver AAT and *Synechocystis* (Q55128): Synech AAT. (Q02635), *Rhizobium meliloti*: R. meli AAT (P58350), *Streptomyces avermitilis*: S. aver AAT (Q82DR2) and *Synechocystis*: Synech AAT (Q55128).

A

```
Sco_Q3AK33    MPRPPELSDRAVALKPSLTLEISAKAKALRDSGKDICSLSAGEPDFATPPFIVEAAQHALGSGFTRY
Syn_Q55128    ----MRLTQRVSQVVPSITLEITAKAKAMRTEGIDVLSFTAGEPDFTTPPHIVEAAKLALDEGKTRY
                   .*::*.  : **:****:*****:* .* *: *::******:***.*****: **..* ***
Sco_Q3AK33    GPAAGDPDLRAALAHKLSVENGIPTQPEQVLVTNGGKQAIYNLFQVLLNPGDEVLLPAPYWLSYPEM
Syn_Q55128    GPAAGEPALRQAIAKKLREKNNLPYEAANILVTNGGKHSLFNLMLAMIEQGDEVIIPAPYWLSYPEM
              *****:* ** *:*:**  :*.:* :  ::*******::::**: .::: ****::***********.
```

B



C

	Sco_Q3AK33	
Substrate varied/fixed	K_M (µM)	K_{cat}(s^{-1})
Oxaloacetate/Glu	12±2	16±1,8
Prephenate/Glu	207±21	0,25±0,05

Figure 5: Analysis of recombinant *Synechochoccus* Q3AK33 1βAAT. A: N-terminal sequence alignment between *Synechochoccus* Q3AK33 and *Synechocystis* Q55128 1βAAT. B: SDS-PAGE of recombinante Sco_Q3AK33 Q sepharose purification: 1 total protein extract 20 µg; 2: Soluble protein extract 20 µg; 3 most active fraction eluted from a Qsepharose column 10 µg. C: Kinetic characterization of *Synechochoccus* Q3AK33 1βAAT. Glutamate 25mM was used as amino donor**.** Values given are the average of at least three independent determinations from a single enzyme preparation ± SD.

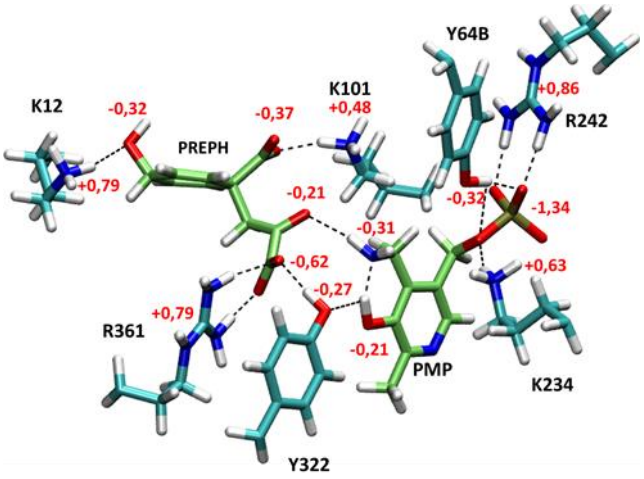


Figure 6: View of the QM active site around prephenate and PMP after QM/MM optimization with Gaussian/Tinker. QM waters, W125, S127 and MM atoms are omitted for clarity. Hydrogen bonds between protein and ligands are indicated by black dashes. Partial charges of the different functional groups are indicated (PREPH, PMP and Tyr hydroxyls, PREPH carboxylates and carbonyl, Lys ammoniums, Arg guadinidium).

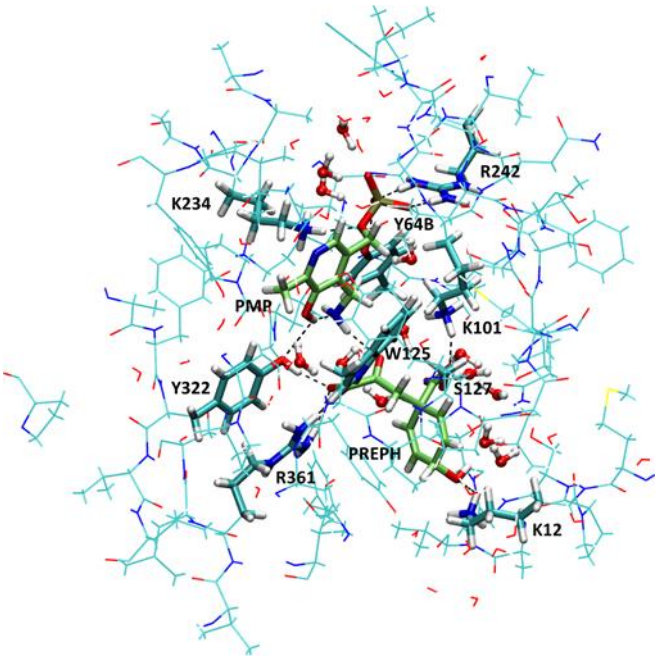


Figure 7: **View of QM and QM/MM parts of the QM/MM simulation in an orientation close to Fig. 3**. The QM part (shown in licorice) includes several residues around the active site with PREPH and PMP (shown in Fig. 5) plus W125 and S127 side chains and 14 water molecules shown in balls and sticks. All other represented atoms are MM atoms belonging to residues within a distance of 3 Å of the QM part and allowed to relax during geometry optimizations, while all other atoms of the enzyme were kept frozen.

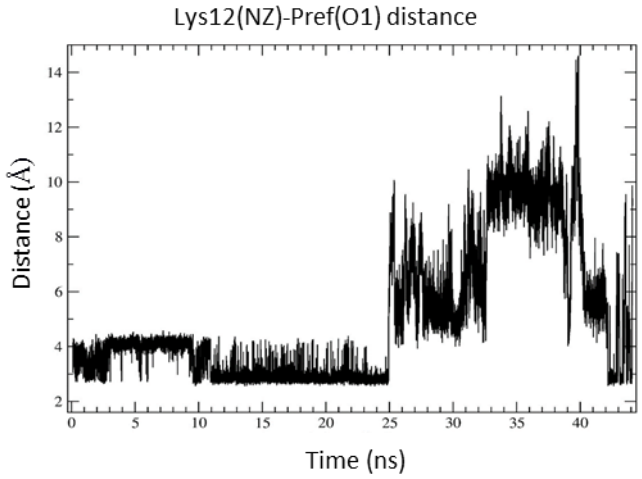


Figure8: Evolution of dK-Prf vs. time during a constant Pressure and Temperature dynamics. The simulation starts with the old set of charges and NOE restraints until 20 ns (S0). Then, charges were modified according to QM/MM parameterization and the simulation was restarted for 4 ns. Finally, all NOE restraints were removed for the remaining 20 ns (24 to 44 ns).

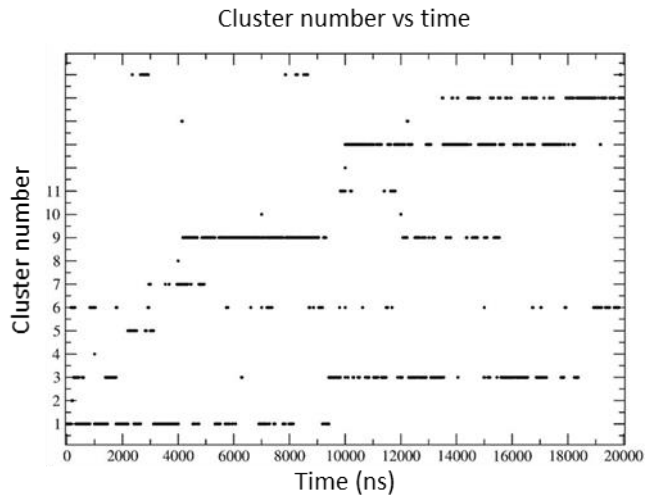


Figure 9: Evolution of cluster population versus time, following 16 20 ns RXSGLD simulations. Some clusters like cluster 1 or 7 appear at the beginning of the simulation and then disappear while others show up after 14 ns like cluster 15. Others like 2, 4, 8, 10, 12 or 14 appear only transiently. Clusters 1, 3, 6, 9, 13 and 15 are the most populated (>= 5%).

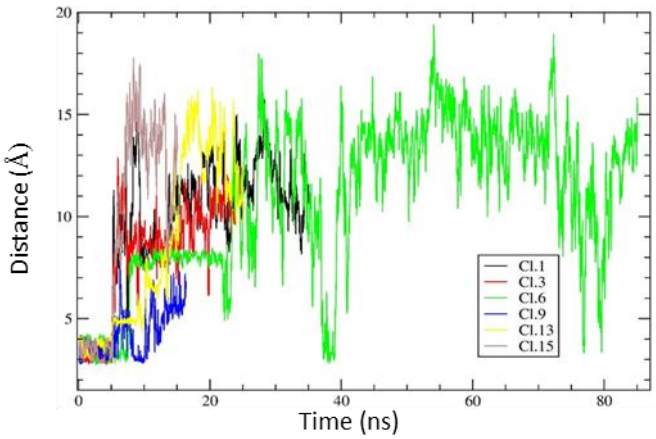


Figure 10: Evolution of dK-Prf vs. time during constant Pressure and Temperature dynamics simulations on selected clusters. The Noe restraints were removed after 5 ns. (Traces were smoothed by moving averages over a 20 point window). Most simulations show a rapid increase of dK-Prf just after the release of the restraints (5 ns). Although longer simulation times would be necessary to follow each of them, cluster 6 shows a decrease of dK-Prf down to less than 4 Å, already after 22 ns. During the 85 ns simulation of this cluster dK-Prf presents very large fluctuations from 2.6 to more than 21.6 Å

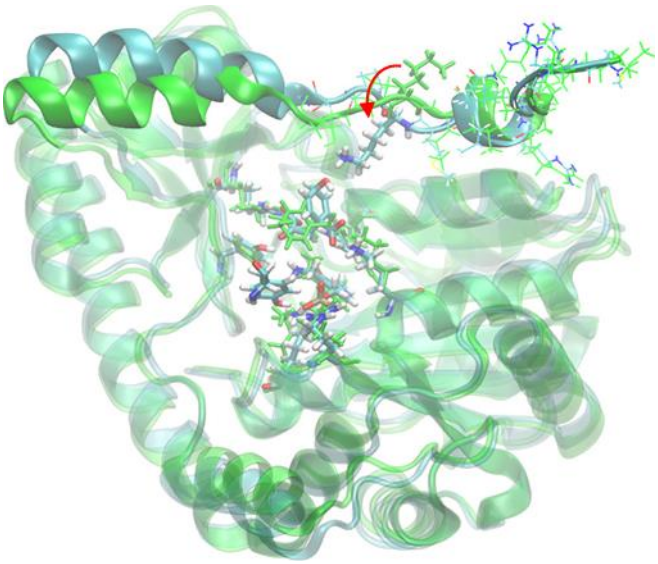


Figure 11: Superimposition of the two structures with largest difference of dK-Prf (Lys12 NZ / prephenate O1 oxygen distance) in the RXSGLD simulation of cluster 6. Structure with dK-Prf= 2.6 Å is shown in cyan and structure with dK-Prf= 21.6 Å is shown in green. The conformational change of Lys12 is clearly visible (red arrow).

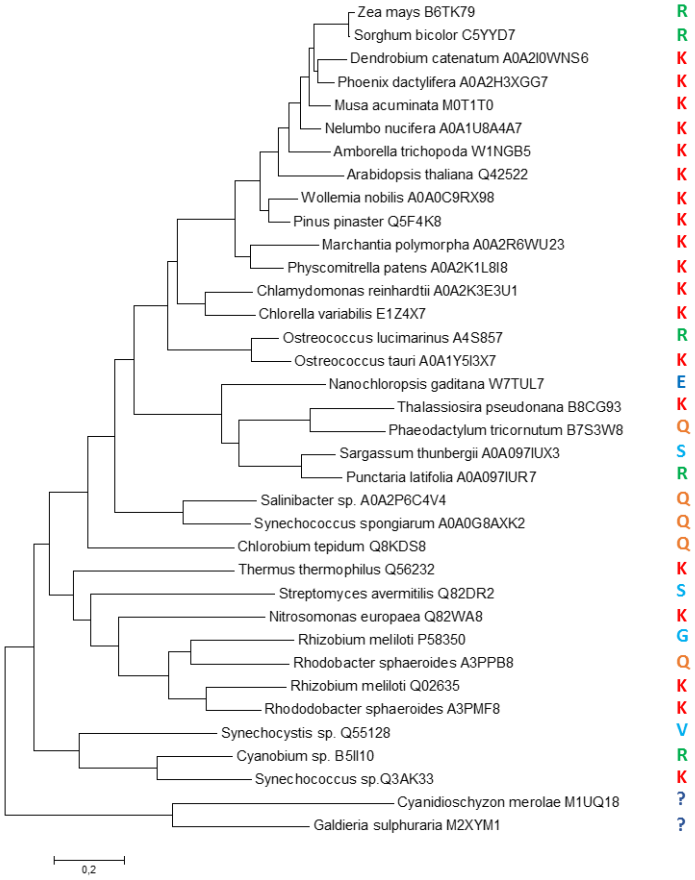


Figure 12: Molecular Phylogenetic analysis of 1βAAT by Maximum Likelihood method
The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-16448,6813) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. The tree is drawn to scale, with branch lengths measured in the

number of substitutions per site. The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 361 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Nature of the amino acid residue in position 12 (according to Thermus thermophilus 1βAAT) is represented.

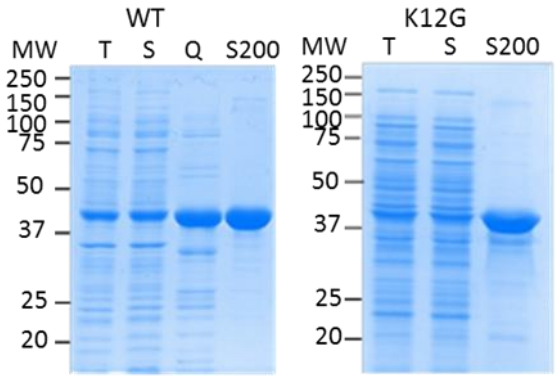


Figure 13: SDSPAGE of the partial purification of recombinant *Rhizobium meliloti* 1βAAT PAT (Q02635) wild type and K12G mutant.
MW: molecular weight; T: total protein extract (20µg); S: soluble protein extract (20µg); Q: Q sepharose pool (10µg); S200: S200 pool (10µg).

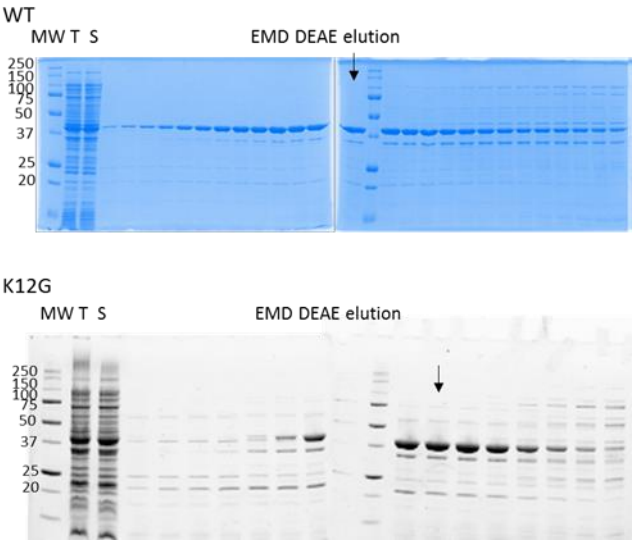


Figure 14: SDS-PAGE of EMD DEAE partial purification of recombinant *Thermus thermophilus* 1βAAT Q56232 wild type and K12G mutant: T: total protein extract (20 µg); S: Soluble protein extract (20 µg). Arrows indicate the most active fraction used for kinetic analysis.