

OPEN

Exploring the therapeutic potential of modern and ancestral phenylalanine/tyrosine ammonia-lyases as supplementary treatment of hereditary tyrosinemia

Natalie M. Hendrikse^{1,2,3}, Albin Holmberg Larsson^{1,2}, Stefan Svensson Gelius¹, Sergei Kuprin¹, Erik Nordling¹ & Per-Olof Syrén^{2,3,4*}

Phenylalanine/tyrosine ammonia-lyases (PAL/TALs) have been approved by the FDA for treatment of phenylketonuria and may harbour potential for complementary treatment of hereditary tyrosinemia Type I. Herein, we explore ancestral sequence reconstruction as an enzyme engineering tool to enhance the therapeutic potential of PAL/TALs. We reconstructed putative ancestors from fungi and compared their catalytic activity and stability to two modern fungal PAL/TALs. Surprisingly, most putative ancestors could be expressed as functional tetramers in *Escherichia coli* and thus retained their ability to oligomerize. All ancestral enzymes displayed increased thermostability compared to both modern enzymes, however, the increase in thermostability was accompanied by a loss in catalytic turnover. One reconstructed ancestral enzyme in particular could be interesting for further drug development, as its ratio of specific activities is more favourable towards tyrosine and it is more thermostable than both modern enzymes. Moreover, long-term stability assessment showed that this variant retained substantially more activity after prolonged incubation at 25 °C and 37 °C, as well as an increased resistance to incubation at 60 °C. Both of these factors are indicative of an extended shelf-life of biopharmaceuticals. We believe that ancestral sequence reconstruction has potential for enhancing the properties of enzyme therapeutics, especially with respect to stability. This work further illustrates that resurrection of putative ancestral oligomeric proteins is feasible and provides insight into the extent of conservation of a functional oligomerization surface area from ancestor to modern enzyme.

Hereditary tyrosinemia (HT) is the collective name of rare hereditary congenital diseases that are characterised by an impaired breakdown of L-tyrosine (L-Tyr). Three types of tyrosinemia are defined to date, which are associated with impaired or depleted function of three distinct enzymes in the breakdown pathway¹ (Supplementary Fig. S1). In HT Type I (HT1), mutations in the fumarylacetoacetate hydrolase gene result in loss of function of the corresponding enzyme (EC 3.7.1.2), effectively blocking the final step of L-Tyr catabolism. This in turn leads to the accumulation of toxic intermediates maleylacetoacetate and fumarylacetoacetate in body fluids and organs, which can induce organ failure and carcinogenesis. HT1 patients are currently treated with Nitisinone (NTBC), which inhibits an upstream enzyme (Supplementary Fig. S1) and thereby prevents the accumulation of toxic intermediates, combined with a protein-restricted diet². Since NTBC terminates the pathway at an earlier stage—corresponding to the phenotype of HT Type III—L-Tyr and 4-hydroxyphenylpyruvate accumulate in plasma. It has been observed that HT1 patients treated with NTBC and a protein restricted diet are at risk of presenting

¹Swedish Orphan Biovitrum AB, SE-112 76, Stockholm, Sweden. ²School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Science for Life Laboratory, Tomtebodavägen 23, Box 1031, SE-171 21, Solna, Sweden. ³School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Fibre and Polymer Technology, KTH Royal Institute of Technology, Teknikringen 56-58, SE-100 44, Stockholm, Sweden. ⁴School of Engineering Sciences in Chemistry, Biotechnology and Health, Division of Protein Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Tomtebodavägen 23, Box 1031, SE-171 21, Solna, Sweden. *email: per-olof.syren@biotech.kth.se

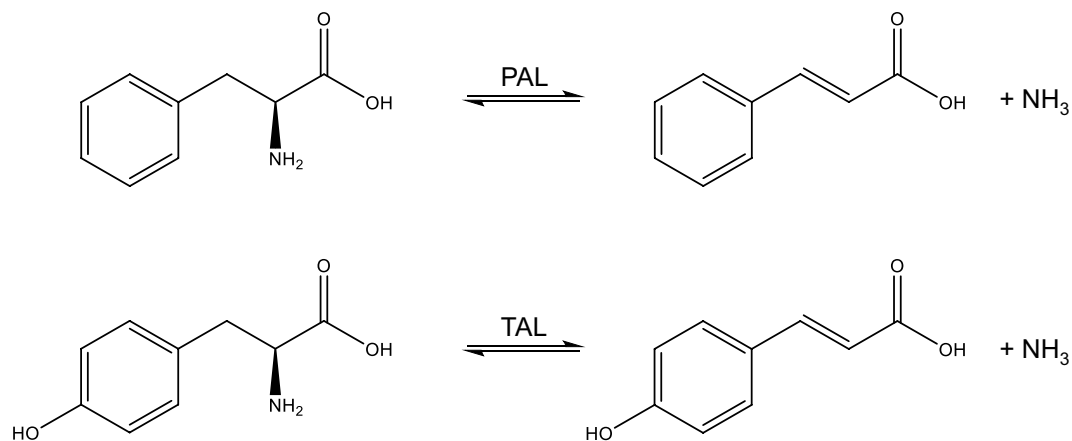


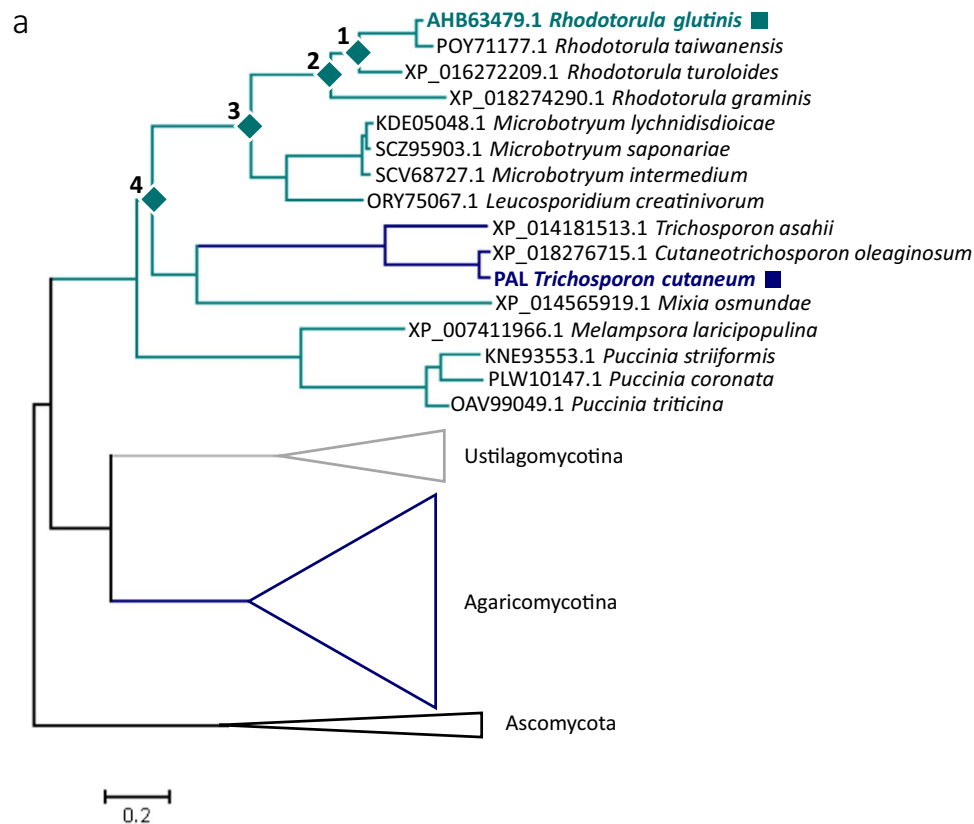
Figure 1. PAL and TAL catalyse the deamination of L-phenylalanine to cinnamic acid and ammonia and L-tyrosine to coumaric acid and ammonia, respectively. Charges are not shown.

developmental delay and impaired cognitive functioning, possibly due to the elevated concentration of L-Tyr³. Moreover, L-Phenylalanine (L-Phe) accumulates in blood plasma, as its catabolism normally proceeds *via* the L-Tyr breakdown pathway (Supplementary Fig. S1). Excess L-Phe, or hyperphenylalaninemia, is also damaging to tissues and has been associated with irreversible intellectual and developmental disabilities in connection with phenylketonuria (PKU); the genetic disease associated with impaired L-Phe breakdown^{4–6}. In order to bypass the natural L-Phe/L-Tyr breakdown pathway and avoid accumulation of toxic or otherwise harmful metabolites, alternative pathways can be explored, which has previously been done for PKU^{7–10}.

Tyrosine ammonia-lyase (TAL, EC 4.3.1.23) is an enzyme that catalyses the non-oxidative deamination of L-Tyr into *p*-coumaric acid (Fig. 1) and was first identified in bacteria¹¹. TAL is part of the aromatic amino acid ammonia-lyase family that also includes phenylalanine ammonia-lyase (PAL, EC 4.3.1.24), histidine ammonia-lyase (HAL, EC 4.3.1.3) and structurally and mechanistically similar class I lyase-like aminomutases¹². The enzymes in this family are characterised by conserved residues Ala-Ser-Gly in the active site that autocatalytically form the activated group 4-methylideneimidazole-5-one¹³ (MIO) to generate a catalytically competent enzyme that is functional as a homotetramer¹⁴. The biocatalytic potential of this enzyme family has been explored for several applications^{12,15,16}. Since both *p*-coumaric acid (Fig. 1, bottom) and cinnamic acid (Fig. 1, top) are non-toxic metabolites that are eventually excreted in urine^{17,18}, TAL and PAL are potentially interesting therapeutic enzymes for treatment of HT1. In analogy, subcutaneous injection of a prokaryotic PEGylated PAL from *Anabaena variabilis* was approved by the FDA in 2018 for the treatment of PKU¹⁹.

Over the last two decades, several PAL enzymes that also have side activity towards L-Tyr have been identified, mostly from fungi and monocotyledonous plants^{20,21}. These enzymes with dual activity are often referred to as PAL/TAL enzymes (EC 4.3.1.25) and some deaminate both substrates with catalytic turnovers of >1 per second²², which is higher than most reported activities for L-Tyr-specific TAL from bacteria (typical catalytic turnover ca. 0.1 s⁻¹)²³. Therefore, we believe that this PAL/TAL group harbours potential therapeutic enzymes that could be used for complementary treatment of HT1 in addition to NTBC, relieving the burden of a strict diet for patients as well as reducing accumulation of both L-Tyr and L-Phe. In order to enhance the therapeutic potential beyond existing PAL/TALs, in particular in terms of protein stability and relaxed substrate specificity, we turned our attention towards ancestral sequence reconstruction (ASR) as a strategy for enzyme engineering^{24–26}. The benefit of using this method for enhancing properties that are beneficial for biopharmaceuticals has previously been shown for coagulation factor VIII²⁴. In our group we previously explored this approach for a terpene cyclase enzyme and obtained enzyme scaffolds with increased thermostability and promiscuity²⁷. Both of these properties would be of interest for a therapeutic PAL/TAL, for which prolonged half-life in the body and ability to deaminate both L-Phe and L-Tyr are of importance. Ancestral sequence reconstruction of an oligomeric protein (in this case a homotetramer) would provide insight into the performance of this method when applied to multimeric targets, a challenge which hitherto has been scarcely explored^{28,29}. Moreover, our approach would also shed light on fundamental aspects of protein oligomerization that have attracted scientific interest^{30,31}.

As a starting point we selected two fungal enzymes that are both reported to have relatively high L-Tyr activity: PAL/TAL from *Rhodotorula glutinis* (RgPAL) and PAL/TAL from *Trichosporon cutaneum* (TcPAL)^{21,22}. We reconstructed several ancestral variants within their clade of the tree and evaluated their therapeutic potential in terms of activity and stability. For comparison, an extended characterization of the two modern enzymes was performed. We found that all ancestral variants had increased thermostability compared to RgPAL and TcPAL, albeit at the cost of lower catalytic turnover. Moreover, one ancestral variant showed considerably improved long-term stability as well as increased ability to maintain activity after incubation at 60 °C, which are both indicative of a potentially improved shelf-life. All ancestors retained the capability to oligomerize, supporting an evolutionary model that does not pass through an intermediate state with disrupted protein-protein interactions³⁰. We believe that our results highlight the potential of ancestral sequence reconstruction as an approach to engineer multimeric enzymes towards therapeutic applications, especially with respect to stability.



b

	RgPAL	MEGA_A1	MEGA_A3	PAML_A1	PAML_A2	PAML_A3	PAML_A4
RgPAL	100.0	85.8	72.9	87.2	82.1	73.8	63.6
MEGA_A1	85.8	100.0	78.6	94.6	89.4	79.4	67.7
MEGA_A3	72.9	78.6	100.0	77.6	81.9	89.7	77.8

Figure 2. Ancestral sequence reconstruction of PAL/TAL. (a) Maximum Likelihood tree of PAL/TALs from fungi. The three subphyla from Basidiomycota are colored blue (Agaricomycotina), grey (Ustilagomycotina) and green (Pucciniomycotina). Sequences from Ascomycota were used as an outgroup (black). All clades apart from the Pucciniomycotina are shown as triangles, the full tree can be found in Supplementary Fig. S2. Ancestral nodes are indicated by diamonds and the two query sequences are indicated by squares. (b) Sequence identities shared by RgPAL and the reconstructed ancestral sequences.

Results

Reconstruction of ancestral PAL/TALs from fungi by MEGA and PAML. Due to their relatively favourable L-Phe/L-Tyr specificity ratio, RgPAL and TcPAL were selected as query sequences for construction of a phylogenetic tree of 68 fungal PAL/TALs (Fig. 2a, full tree in Supplementary Fig. S2). The two sequences share 50% sequence identity between them and have reported PAL/TAL specificity ratios of 2–3^{22,32}. Remarkably, three sequences from subphylum Agaricomycotina—amongst which the sequence from *T. cutaneum*—cluster with the Pucciniomycotina sequences. Apart from this exception the tree topology roughly follows a species tree of Basidiomycota that was previously published³³. We used maximum likelihood statistics implemented in PAML^{34,35} to reconstruct the most likely ancestral sequences at four different nodes (Fig. 2a). Going from young to old, A1, A2, A3 and A4 roughly span the evolutionary distance between RgPAL to the last common ancestral node with TcPAL. Two of the four nodes were also chosen for ancestral sequence reconstruction in MEGA⁷³⁶ to increase our coverage of ancestral sequence space and to investigate the robustness of the reconstruction. The putative ancestral sequences differ from RgPAL by 13–34% in sequence identity (Fig. 2b). Figure 2b also shows that the respective sequences of A1 and A3 from PAML and MEGA differ slightly in identity compared to RgPAL: the PAML sequences are ca. 1% closer to RgPAL. The two variants of the same node share sequence identities of 95% (A1) and 90% (A3), including some small indels. Inspection of the reconstructed ancestral sequences confirmed that all residues known to be catalytically important were conserved, including the Ala-Ser-Gly triad in the active site that forms the catalytic moiety MIO (Supplementary Fig. S3).

Homology model of MEGA_A1 reveals that oligomerization interface is not conserved. As the modern enzymes are active as homotetramers²¹, we were interested to see if and how the interfaces between monomers changed in the reconstruction process. To visualize the spatial distribution of ancestral mutations, a

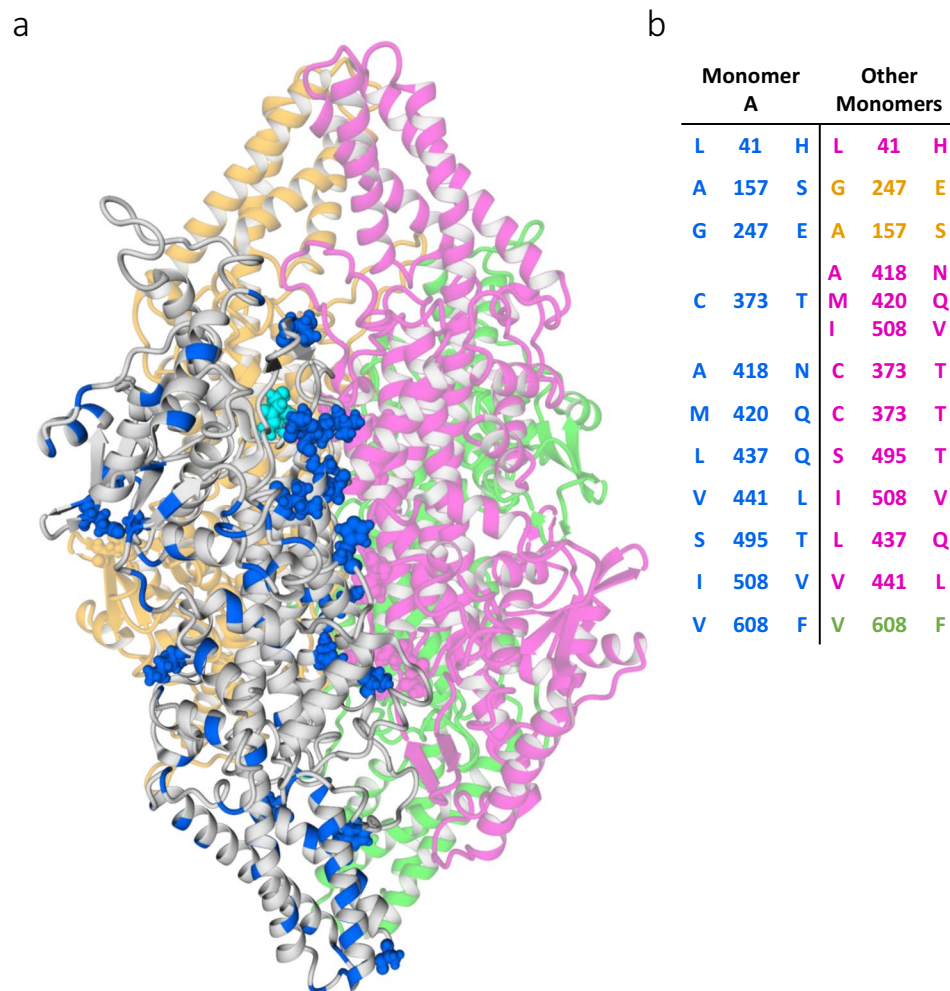


Figure 3. Ancestral PAL/TALs form tetramers despite non-conserved interfaces between monomers. **(a)** Homology model of MEGA_A1 showing the distribution of ancestral mutations. Monomers are coloured in grey (monomer A), magenta (monomer B), yellow (monomer C) and green (monomer D). All ancestral mutations are coloured in blue in monomer A, and surface mutations within 4 Å of another monomer are shown as balls. Catalytically active group MIO is shown in cyan for monomer A for reference of the active site location. The model was built in YASARA using the crystal structure of PAL/TAL from *Rhodotorula turoloides* (PDB, 1Y2M⁷) as a template. **(b)** Possibly coupled ancestral mutations in the oligomerization interfaces. The mutations for monomer A are listed in blue and the corresponding mutations in proximity are coloured according to monomer. Mutations are noted with RgPAL residues on the left and the corresponding residue in MEGA_A1 on the right.

homology model was built for MEGA_A1 (Fig. 3a). We found that approximately 60% of all ancestral mutations occurred on the surface of the enzyme and that the interfaces between the monomers were not more conserved than other areas on the surface. We noted that some of the mutations in one of the monomers in the interaction surface were in proximity to ancestral mutations in the other monomer. In fact, eleven interaction surface mutations from monomer A were within 4 Å of a mutation in another monomer (Fig. 3b).

Ancestral enzymes can be functionally expressed in *E. coli*. All enzymes except PAML_A4 could be successfully expressed in *E. coli* BL21 (DE3) and were purified by affinity chromatography followed by size exclusion chromatography. During the second step of the purification process we noted that some enzymes eluted from the size exclusion column in three separate peaks. As the active form of the enzymes has previously been found to be a homotetramer²¹, we hypothesized that these peaks may constitute different oligomeric states. To investigate this hypothesis, we applied analytical size exclusion chromatography (Fig. 4a) coupled to a multi-angle light scattering detector (SEC-MALS) to determine the approximate molecular weight of each peak eluting from the column (Fig. 4b). The results indicated that the tetramer was in fact the smallest oligomer eluting from the SEC column and that the two other peaks consisted of larger aggregates.

Interestingly, the ratio between the areas of the different peaks was not the same for all enzymes: whereas RgPAL and PAML_A3 had significant peaks with higher molecular weight, TcPAL, MEGA_A1, MEGA_A3, PAML_A1 and PAML_A2 appeared mainly in tetrameric form. An initial activity screen of 15 minutes was

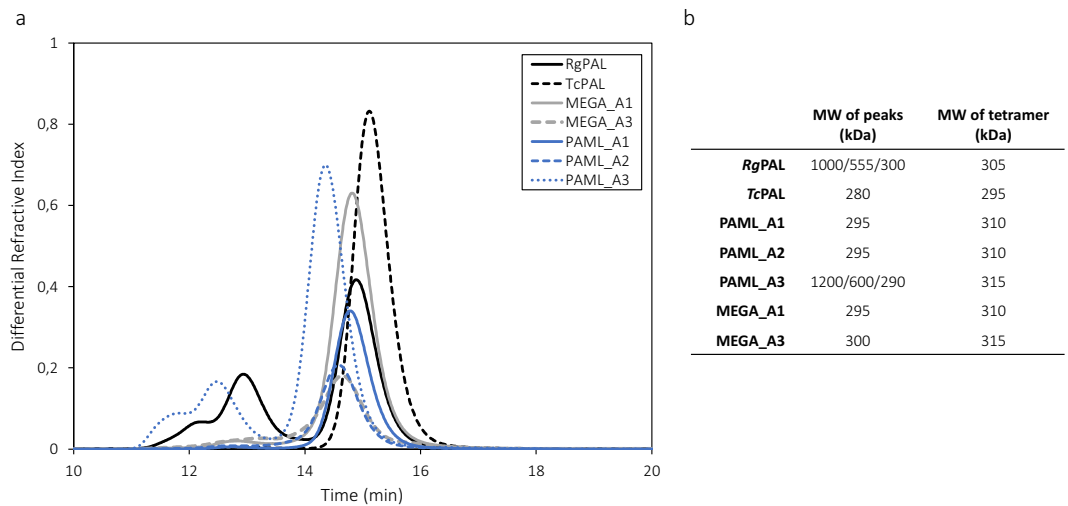


Figure 4. Analysis of oligomers by SEC-MALS. **(a)** Elution profiles of all enzymes from the size-exclusion chromatography column. **(b)** Molecular weights (MW) of the SEC-elution peaks as determined by MALS (left column). For comparison, the theoretically calculated MW for the respective tetramers is shown the right column. For *RgPAL* and *PAML_A3*, MW is shown for all three peaks.

	L-Phe ^a			L-Tyr ^a			Relative Activity ^b	
	k_{cat} (s ⁻¹)	K_M (mM)	k_{cat}/K_M (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_M (mM)	k_{cat}/K_M (M ⁻¹ s ⁻¹)	$\frac{[k_{cat}]_{Phe}}{[k_{cat}]_{Tyr}}$	$\frac{[k_{cat}/K_M]_{Phe}}{[k_{cat}/K_M]_{Tyr}}$
<i>RgPAL</i>	8.95 ± 0.37	1.35 ± 0.10	6655 ± 259	1.71 ± 0.11	0.28 ± 0.01	6303 ± 288	5.2	1.1
<i>TcPAL</i>	13.24 ± 0.42	0.10 ± 0.03	— ^c	2.21 ± 0.28	0.07 ± 0.05	— ^c	6.0	— ^c
<i>MEGA_A1</i>	5.76 ± 0.25	3.83 ± 0.97	1560 ± 326	1.05 ± 0.06	0.40 ± 0.02	2643 ± 312	5.5	0.6
<i>MEGA_A3</i>	1.42 ± 0.03	0.35 ± 0.03	4069 ± 356	0.23 ± 0.01	0.05 ± 0.01	4509 ± 34	6.2	0.9
<i>PAML_A1</i>	0.30 ± 0.01	4.48 ± 0.62	68 ± 8	0.04 ± 0.01	2.69 ± 0.01	17 ± 1	6.8	4.1
<i>PAML_A2</i>	0.03 ± 0.01	0.69 ± 0.15	42 ± 7	— ^d	— ^d	— ^d	— ^d	— ^d

Table 1. Kinetic parameters of all enzymes for L-Phenylalanine and L-Tyrosine at 37 °C. Standard deviations are shown with $n = 3$. ^aMeasurements were performed in a 50 mM Tris-HCl buffer at pH 9.2. ^bDifferences may occur due to rounding. ^cFor *TcPAL*, K' is shown instead of K_M due to cooperativity. Hill constants were 1.6 ± 0.03 (L-Phe) and 1.5 ± 0.1 (L-Tyr). ^dFor *PAML_A2*, activity towards L-Tyr was not sufficient for kinetic analysis.

performed for all purified enzymes with both L-Phe and L-Tyr as substrates (Supplementary Fig. S4). The highest activity was detected for *RgPAL*, *TcPAL* and *MEGA_A1*, moderate activity was detected for *PAML_A1* and *MEGA_A3*, and low activity for *PAML_A2*. After extended incubation for 1 hour, activity could also be detected for *PAML_A3* (data not shown). Based on these results we decided to perform kinetic analysis for *RgPAL*, *TcPAL*, *PAML_A1*, *PAML_A2*, *MEGA_A1* and *MEGA_A3* to enable quantification and comparison of specific activities of these enzymes.

Kinetic analyses show cooperativity in *TcPAL* and reduced activity in ancestral variants.

Kinetic assays were performed separately with L-Phe and L-Tyr at 37 °C (Supplementary Fig. S5). Kinetic parameters can be found in Table 1. We observed cooperativity in *TcPAL*, characterised by the non-linear relation between reaction rate and substrate concentration at low substrate concentrations, which has also been reported previously for this enzyme^{22,32}. We observed moderate cooperativity with both L-Phe and L-Tyr and determined the Hill constants to be 1.6 ± 0.03 and 1.5 ± 0.1 , respectively. We found that the catalytic turnover of both substrates was higher for *TcPAL* as compared to *RgPAL*, yet the ratio between the PAL/TAL activities was comparable (Table 1). Kinetic assays of the ancestral variants confirmed the general trend observed in the screen: activity was reduced in nearly all ancestral variants and decreased with each step back in the tree. The two ancestors that were reconstructed in *MEGA* had higher activity than those reconstructed in *PAML* and cooperativity was not observed in any of the ancestors. We observed similar PAL/TAL ratios of catalytic turnover for all enzymes, but changes in PAL/TAL specificity ratio could be seen. The specificity ratio for both *MEGA* ancestors became favourable towards L-Tyr, whereas *RgPAL* favoured L-Phe. Interestingly, both older ancestors *MEGA_A3* and *PAML_A2* showed substantially decreased K_M values as compared to their younger counterparts.

	T_m (°C) ^a
TcPAL	61.7 ± 0.1
RgPAL	66.2 ± 0.1
PAML_A1	71.4 ± 0.1
PAML_A2	74.8 ± 0.6
PAML_A3	78.0 ± 0.8
MEGA_A1	70.9 ± 0.1
MEGA_A3	<i>n.d.</i>

Table 2. Analysis of thermostability by nanoDSF. Standard deviations are shown with $n = 3$. ^aMeasurements were performed in a 50 mM Tris-HCl, 125 mM NaCl buffer, pH 8.5.

Analysis of thermostability shows increased stability for all ancestors. In order to characterise the thermostability of the ancestral PAL/TAL variants we established melting curves using nanoDSF^{37,38} followed by determination of melting temperatures (Table 2, Supplementary Fig. S6). Modern enzymes TcPAL and RgPAL had the lowest melting temperatures of 62 °C and 66 °C, respectively. For the three ancestors reconstructed in PAML, stability increased with each step back in the tree: 71 °C (PAML_A1), 75 °C (PAML_A2), and 79 °C (PAML_A3). The melting temperature of MEGA_A1 was determined to be 71 °C, but for MEGA_A3 no clear shift in tryptophan emission wavelength could be seen (data not shown). A look at the tryptophan content of all enzymes revealed that RgPAL and all PAML ancestors share the same four tryptophans, three of which also occur in the sequence of TcPAL (Supplementary Fig. S3). MEGA_A1 and MEGA_A3 have 5 and 6 tryptophans, respectively, of which one in MEGA_A3 is located close to the N-terminus and may mask the signal of the others during thermal unfolding. To obtain information on the relative thermostability of MEGA_A3, thermal unfolding of the two modern enzymes and the two MEGA ancestors was also measured using CD spectroscopy (Supplementary Fig. S7). CD spectra showed that loss of secondary structure of MEGA_A1 and MEGA_A3 occurred at higher temperatures compared to both modern enzymes.

Evaluation of therapeutic potential highlights increased stability of MEGA_A1 and MEGA_A3.

To investigate the potential of oral administration, proteolytic assays with two different proteases were performed for RgPAL and MEGA_A1. We chose to include pepsin and trypsin; digestive enzymes from the stomach and duodenum, respectively. The reactions with pepsin were carried out at its optimum pH of 2.0 and the reactions with trypsin were carried out at pH 8.5, which is close to the optimum pH for the lyase enzymes³⁹. Samples were taken at various timepoints and analysed on SDS-PAGE in addition to activity measurements with L-Tyr. Both the gel and the activity measurements showed that RgPAL and MEGA_A1 were immediately inactivated by pepsin (Supplementary Fig. S8). Interestingly, the control samples without pepsin did not show enzymatic activity even though the pH had been restored to 8.5 for the activity measurements. Samples with and without trypsin were monitored over the course of 4 hours (Supplementary Fig. S8). After 4 hours both RgPAL and MEGA_A1 had lost approximately 55% of their activity (Supplementary Fig. S8b) and an additional measurement after 20 hours confirmed that activity had completely disappeared. Another possible way of administration is intravenously, in which case the enzymes would need to be active in the blood. To explore this possibility, we aimed to confirm activity of MEGA_A1 and RgPAL in a more physiologically relevant matrix. Calf serum was spiked with 800 μM L-Tyr to mimic levels in patients¹ and a range of enzyme concentrations was tested to get an estimate of an appropriate dose (Fig. 5). No L-Phe was added, but approximately 200 μM was present in the serum initially. All L-Phe was cleared from samples with 2000 nM RgPAL and MEGA_A1, whereas L-Tyr was reduced by 95% and 83%, respectively.

To further explore their potential as therapeutics, samples of RgPAL, TcPAL and MEGA_A1 were incubated at 25 °C and 37 °C for 2–4 weeks, followed by evaluation of stability as well as L-Phe activity (Fig. 6a,b). Figure 6a shows that MEGA_A1 samples were most active in all three cases and that up to 52% of activity was retained after 4 weeks at 37 °C. Even TcPAL was significantly more stable than RgPAL, which lost activity completely after 2 weeks at 37 °C. Analysis on SDS-PAGE (Fig. 6b) supports these findings as the band around 75 kDa becomes fainter in RgPAL samples and slightly fainter in TcPAL samples, but not for MEGA_A1. All three enzymes seemed more tolerant to 4 weeks at 25 °C than 2 weeks at 37 °C. We were also interested in testing heat resistance of the most active ancestral variants, MEGA_A1 and MEGA_A3 in comparison to RgPAL, the modern variant displaying the highest T_m (Table 2). Figure 6c shows the relative activity of the three enzymes after incubation at 60 °C for 30 and 60 minutes, as compared to a control not subjected to heat. Whereas RgPAL lost nearly all activity after 60 minutes, MEGA_A1 and MEGA_A3 retained 63% and 64% of their activity, respectively.

Discussion

In this study we aimed to explore if ancestral sequence reconstruction could be a viable strategy to enhance therapeutic properties of phenylalanine/tyrosine ammonia-lyases from fungi, towards a complementary treatment of hereditary tyrosinemia. We found that all ancestral variants showed increased thermostability, albeit at the expense of catalytic turnover as compared to modern enzymes RgPAL and TcPAL. MEGA_A1 displayed substantially enhanced long-term stability at 25 and 37 °C and, together with MEGA_A3, significantly increased resistance to incubation at 60 °C compared to RgPAL, which highlights the possibilities of ancestral sequence reconstruction for engineering stable oligomeric therapeutic enzymes.

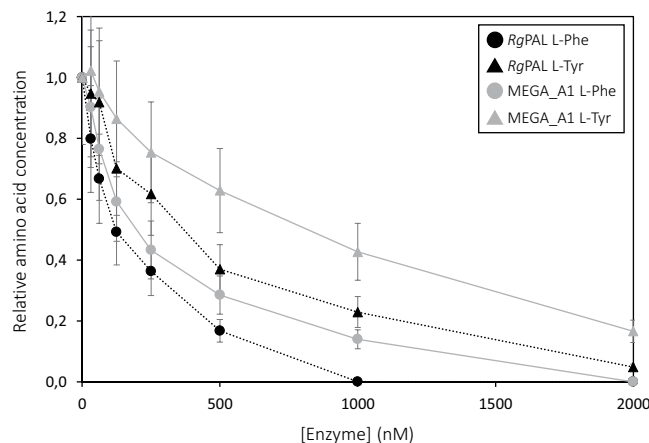


Figure 5. Activity of RgPAL and MEGA_A1 in calf serum. Relative amino acid concentrations are shown for different enzyme concentrations. Starting concentrations were approximately 1000 μM for L-Tyr and 200 μM for L-Phe and final concentrations were determined by reversed-phase HPLC, after incubation with enzyme. Error bars shown are based on duplicates of the $[E] = 0$ measurement.

It has previously been reported that fungal PAL is monophyletic, *i.e.* that all genes derive from a common ancestral gene that was present in the ancestor of the subkingdom Dikarya or earlier⁴⁰. We used two different programs—MEGA and PAML—for ancestral reconstruction of PAL/TALs at several nodes within the Pucciniomycotina, a subphylum of the division Basidiomycota. The divergence of the three subphyla in this division is estimated to have occurred between 1030 and 670 million years ago, but remains ambiguous⁴¹, which makes it difficult to estimate an age for the ancestral enzymes. Interestingly, a group of three sequences from the Agaricomycotina was placed within the clade of Pucciniomycotina with high support values, in trees built with Maximum Likelihood as well as Bayesian Inference (Supplementary Figs. S2 and S9). Amongst these was the query sequence from *T. cutaneum* that, in contrast to the other characterised enzymes, displays cooperativity in kinetic analyses^{22,32}. This tree topology and the fact that it is the only enzyme displaying cooperative kinetics could indicate that a horizontal gene transfer took place between the two subphyla, which would mean that node 4 is not a true divergence point in evolution. In that case node 4 does not represent an actual common ancestor of the sequences in that clade, which could be a possible explanation as to why PAML_A4 cannot be functionally expressed⁴². In this study we used two different methods to reconstruct ancestral sequences that are both based on maximum likelihood statistics^{34,36}. Despite the relatively small differences between the sequences from MEGA and PAML—5% at node 1 and 10% at node 3—differences in activity of up to two orders of magnitude were observed. As all residues known to be catalytically relevant are conserved, there is no obvious sequence-based explanation for the lower activity of the PAML variants. Most differences, including some small insertions in both sequences, occur close to the N-terminal, which is highly variable in all sequences (Supplementary Fig. S3). Moreover, the knowledge regarding the catalytic mechanism of the enzyme family remains ambiguous^{12,43,44}, making it highly challenging to understand the influence of point mutations from a structural perspective.

Apart from the generally reduced activity of the ancestral enzymes, we observed a decrease in k_{cat} accompanied by a decrease in K_{M} for both MEGA and PAML ancestors when going back in the phylogenetic tree. This finding could be in accordance with the previously proposed stability-activity trade-off⁴⁵. As we envisioned the use of this enzyme family for complementary treatment of tyrosinemia, we were particularly interested in turnover of L-Tyr and the ratio between PAL/TAL activity. In terms of k_{cat} we found similar ratios for all enzymes, namely between 5.2 and 6.8, with the lowest being RgPAL followed by MEGA_A1 (5.5). Looking at the catalytic efficiencies, Table 1 shows that both MEGA ancestors have a favourable specificity towards L-Tyr, whereas RgPAL favours L-Phe. K_{M} values for L-Tyr are around 300–400 μM for RgPAL and MEGA_A1 and K' for TcPAL is 70 μM . These values are below typical L-Tyr blood levels in patients with tyrosinemia¹, which means that a therapeutic enzyme would be active under k_{cat} -conditions. In terms of k_{cat} for L-Tyr, RgPAL and TcPAL have values of around 2 s^{-1} , whereas MEGA_A1 has a k_{cat} of 1 s^{-1} . This difference of approximately 1 s^{-1} is less than the absolute differences for L-Phe turnover and its effect on dosage in patients could be compensated for by the increase in long-term stability. Activity measurements in plasma for RgPAL and MEGA_A1 confirm that both enzymes successfully lower L-Tyr levels, despite the sub-optimal pH. Interestingly, MEGA_A3 has a slightly less favourable relative specificity towards L-Tyr than MEGA_A1. However, both modern enzymes in this study are also promiscuous and the evolution of different specificities among the family of aromatic amino acid ammonia-lyases remains unclear⁴⁰.

With respect to the oligomerization state of the enzymes, we found the tetramer to be the main form in solution for all enzymes. Interestingly, analytical SEC elution profiles of RgPAL and PAML_A3 also contained two peaks that seemed to constitute of multitudes of tetramers (2–4) based on the estimated molecular weight (Fig. 4). Similar peak distributions were observed in the SEC purification profile (data not shown). The fact that the ancestral mutations are scattered over the enzyme (Fig. 3a) as well as the lack of knowledge about where on their surface the tetramers interact makes it difficult to pinpoint which residues could influence the interaction.

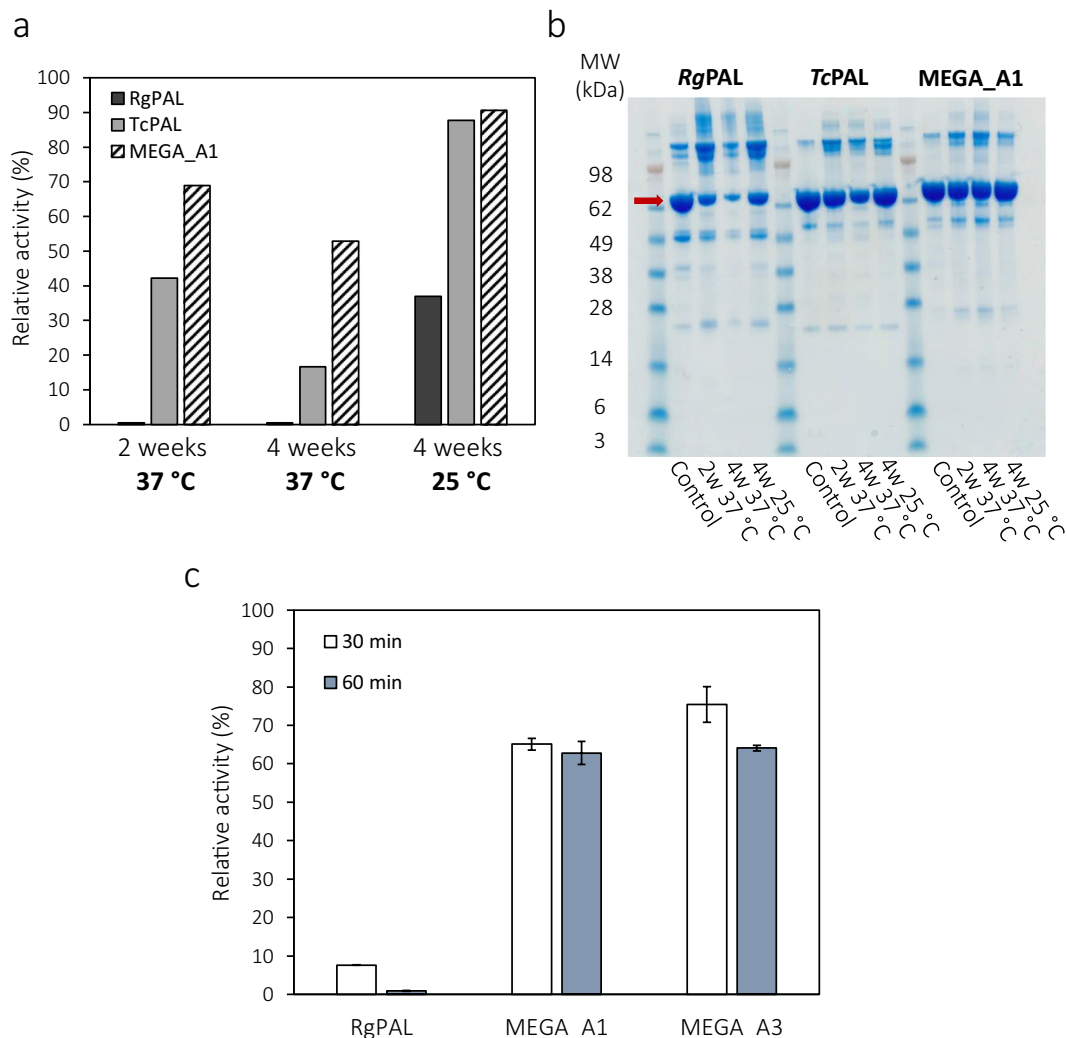


Figure 6. Evaluation of long-term stability. **(a)** Relative L-Phe activity of all long-term stability samples for RgPAL, TcPAL and MEGA_A1. Activity was measured at 37 °C and normalized to control samples that had been stored at -80°C . L-Phe was chosen for activity evaluation based on the higher measured activity towards this substrate (Table 1). **(b)** SDS-PAGE of all samples. The SeeBlue Plus2 pre-stained molecular weight marker was used for reference and PAL enzymes are indicated by an arrow. The gel image was cropped around the edges for display. **(c)** Relative L-Phe activity of RgPAL, MEGA_A1 and MEGA_A3 after incubation at 60 °C for 30 and 60 minutes. Activity was measured at 37 °C and normalized to control samples that had been stored at 25 °C. L-Phe was chosen for activity evaluation based on the higher measured activity towards this substrate (Table 1). Standard deviations are shown with $n = 3$.

Moreover, a large percentage of the mutations in PAML_A3 are shared with PAML_A2 and PAML_A1, both of which only display one major peak in the SEC elution profile. In the context of possible therapeutic applications of these enzymes, the tendency of tetramers to aggregate is clearly a drawback due to an increased risk of immunogenicity⁴⁶. This problem has previously also been described for PAL from *Anabaena variabilis* (the PAL variant approved for treatment of PKU) and was overcome by mutation of two cysteine residues^{47,48}. Another observation regarding the oligomerization was that the interaction interface in MEGA_A1 was not more conserved than the remaining surface area. However, it has previously been shown that oligomeric interaction is not only—or not necessarily—determined by mutations in the actual interface, but also through long-range effects within the subunits^{28,49}. Moreover, eleven ancestral surface mutations in one monomer were accompanied by an ancestral mutation in another monomer within a 4 Å distance (Fig. 3). Some of these mutations could be coupled, which might be accounted for by the reconstruction method since the enzyme is a homotetramer and each subunit contains the same mutations. Formation of larger oligomers than tetramers was reduced for MEGA_A1 compared to RgPAL, which further enhances the therapeutic potential of this variant, despite the decrease in catalytic turnover. Another bioinformatics method often found to increase stability of proteins is the consensus method. However, this approach entails difficulties in identifying correlated mutations as those observed in the interface region⁵⁰. These findings suggest that ancestral sequence reconstruction can also successfully be applied in the context of oligomeric proteins, which is of interest as many biopharmaceutical targets are of complex nature.

In their study, Wang *et al.* also identify the need to improve thermal stability of ammonia-lyase enzymes for therapeutic applications⁴⁸. We found that all ancestral enzymes were more thermostable than RgPAL and TcPAL, with melting temperatures above 70 °C. Perhaps more importantly, MEGA_A1 displayed a substantially increased long-term stability, both at 25 °C and 37 °C. Whereas RgPAL lost all its activity after 2 weeks at 37 °C, MEGA_A1 retained more than 50% of its activity after 4 weeks at 37 °C. In contrast, TcPAL retained 16% of its activity after 4 weeks at 37 °C. These results show a considerable improvement in therapeutic potential for MEGA_A1, as long-term stability directly translates to shelf-life of a drug. Moreover, an increased stability at 37 °C is highly relevant from a therapeutic perspective, as it reflects *in vivo* conditions under which a potential therapeutic protein would act. Molecular mechanisms of thermostability have been discussed in the literature^{51,52}. Overall, the 109 amino acid substitutions in MEGA_A1 compared to RgPAL resulted in an increased number of polar residues, indicating a decrease in hydrophobicity. It is worth to note that the ancestral substitutions suggest a decreased flexibility of MEGA_A1, as the number of glycine residues decreased while the proline content increased (Pro:Gly, RgPAL 27:52; MEGA_A1 30:48). We found that MEGA_A1 and MEGA_A3 displayed increased resistance to incubation at 60 °C compared to RgPAL. Whereas RgPAL lost nearly all activity after incubation for 60 minutes, both ancestral variants retained >60% of their activity. The introduction of proline residues into flexible loops has been shown to increase the thermostability of proteins and contribute to backbone rigidity^{53,54}. For MEGA_A1 the number of prolines was increased by 3 compared to RgPAL, 2 of which are located in loop regions and thus might be able to confer the observed stability increase. As for MEGA_A3, the sequence contains 12 more proline residues as compared to RgPAL and a similar reasoning could be applied there.

Several ways of administration of the therapeutic enzyme could be envisioned, such as oral administration, subcutaneous or intravenous injection. Whereas the former would be aimed at clearing L-Phe and L-Tyr in the gastrointestinal tract before it is absorbed, the latter would reduce blood plasma levels of the amino acids. Both RgPAL and MEGA_A1 were susceptible to degradation by pepsin and trypsin, as well as inactivation by low pH (2.0). Moreover, other studies have shown that RgPAL is also highly susceptible to cleavage by other proteases such as chymotrypsin⁵⁵. Given that oral administration would expose the enzyme to all of the above unless some form of coating is applied⁵⁶, it seems likely that higher *in vivo* activity can be achieved through injection. The pH in serum is approximately 7.5, which is slightly below the optimum pH for PAL/TAL³⁹. However, we have shown that both RgPAL and MEGA_A1 are active in calf serum and decrease both L-Phe and L-Tyr levels, indicating the potential of intravenous administration.

In summary, we obtained ancestral PAL/TAL variants with decreased catalytic turnover, but considerably enhanced stability as compared to two modern fungal enzymes, not only with higher melting temperatures but also promising long-term stability at 37 °C and increased resistance to heat inactivation (as compared to RgPAL). Overall, the ancestral variants displayed properties that could make them potential scaffolds amenable for further engineering towards complementary treatment of tyrosinemia. The ancestral enzymes retained their oligomeric properties even though the interface regions were not more conserved than the exposed surface area, indicating that ancestral sequence reconstruction is feasible for multimeric proteins that are of importance from a biopharmaceutical perspective.

Methods

Ancestral sequence reconstruction and homology modeling. A protein BLAST search was performed via NCBI in the non-redundant database using the sequence of PAL from *Rhodotorula glutinis* (GenBank accession no. AHB63479) as a query. A second BLAST search was performed using the sequence of PAL from *Trichosporon cutaneum*, which was taken from patent US007572612B2⁵⁷. The respective 150 closest eukaryotic homologues were selected for further investigation. Duplicates, incomplete sequences, recombinant proteins and mutants were removed, before aligning the sequences using the L-INS-i algorithm in MAFFT version 7^{58,59}. The sequences were trimmed using trimAl⁶⁰ and the implemented “gappout” method. Both IQ-TREE and MEGA found the LG model⁶¹ +F to be the best model for the dataset, including a gamma distribution (4 categories) to model rate variation across sites. A maximum likelihood tree was then constructed in IQ-TREE⁶² with 1000 bootstrap replicates. Based on the topology of the initial tree, the sequence selection was manually adjusted to generate a selection that covered all three phyla within Basidiomycota. In order to assess robustness of the tree, three independent Markov Chain Monte Carlo chains were run in BALi-Phy version 3.3⁶³ with the same set of sequences, yet in full length. BALi-Phy simultaneously estimates the multiple sequence alignment and phylogenetic tree using Bayesian statistics. The consensus tree from the three runs is shown in Supplementary Fig. S9 and shows that only two minor differences occur with respect to the ML tree - all within the clade of Agaricomycotina. Especially the part in which the ancestors were reconstructed has high posterior probabilities. We inferred the most probable ancestral sequences at four different nodes in PAML version 4³⁵, going back from RgPAL to the last node shared with TcPAL. In order to increase our sampling of ancestral sequence space we also inferred the most probable ancestral sequences at node 1 and 3 in MEGA version 7³⁶. The homology model of MEGA_A1 was built in YASARA⁶⁴ (version 18.4) using the crystal structure of PAL/TAL from *Rhodotorula turoloides* (1Y2M⁷) as a template. Loop regions that were missing from the crystal structure were sampled from the PDB for possible conformations and were finally optimized in YASARA using a YASARA force field. The Z-score of the resulting model was -0.495.

Protein expression and purification. All sequences were codon-optimized for expression in *E. coli* and synthesized by GeneArt (ThermoFisher Scientific) with an N-terminal 6xHis-tag for affinity purification. The genes were subcloned into a pET22b-(+) vector (Novagen, Merck Millipore) using the NdeI and XhoI restriction sites and transformed into *E. coli* BL21 (DE3) cells using a heat-shock protocol. A single colony from each construct was inoculated in 10 mL TB medium containing 100 µg/mL ampicillin and incubated at 160 rpm at 37 °C overnight. The overnight cultures were diluted to an OD₆₀₀ of approximately 0.07 in 40 (ancestors and TcPAL) and 400 mL (RgPAL) TB medium containing 100 µg/mL ampicillin, and were incubated at 160 rpm at 37 °C.

The cultures were induced with 0.4 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at $OD_{600} \approx 0.5$ – 0.7 and expression was performed for 22 hours at 160 rpm at 25 °C. Cells were harvested by centrifugation for 15 min (4 °C, 2264 x g) and the supernatant was discarded. For purification, cell pellets were resuspended in 30 ml resuspension buffer (25 mM Tris-HCl, 500 mM NaCl, 20 mM Imidazole, pH 8.5). The mixtures were sonicated on ice for 2×45 s effective time (5 s on 5 s off at 100% amplitude, Sonics & Materials), and centrifuged for 30 min at 4 °C, 48298 x g (Beckman Coulter). The supernatant was filtered through a 0.22 μ m filter, before being loaded onto a 1 ml HisTrapHP Ni-Sepharose column installed in an ÄKTA Explorer (GE Healthcare). Protein was eluted with elution buffer (25 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazole, pH 8.5) and elution fractions were analyzed on SDS-PAGE. Fractions containing the correctly sized protein were pooled, concentrated, and loaded onto a 120 ml HighLoad 16/60 Superdex200 column (GE Healthcare) which was equilibrated with SEC buffer (50 mM Tris-HCl, 125 mM NaCl, pH 8.5). A flow rate of 1.5 ml/min was used. Elution fractions were analyzed on SDS-PAGE, and fractions containing protein of the correct size were pooled, concentrated and stored at -80 °C.

Analytical size exclusion chromatography. All purified enzymes were diluted to final concentrations of ca. 1–2 mg/ml in SEC buffer. Samples of 50 μ l were loaded onto a Superdex200 Increase 10/300 GL column, which was equilibrated with filtered PBS (pH 7.4) and coupled to an ÄKTAmicro (GE Healthcare). Proteins were eluted with a flow rate of 0.75 ml/min using filtered PBS (pH 7.4), and molecular weight of the proteins in the elution peaks was determined through MALS detection, using a miniDAWN detector (Wyatt Technology). The column was calibrated using low and high molecular weight gel filtration calibration kits (GE Healthcare).

Activity assays and kinetic analyses. Kinetics assays were performed in triplicates at 37 °C. Reactions were carried out in a 50 mM Tris-HCl buffer at pH 9.2 in a total volume of 200 μ L. Substrates were obtained in pure form from Sigma-Aldrich, the final substrate concentrations in the kinetics assay were 44 μ M–45 mM (L-Phe) 35 μ M–9 mM (L-Tyr) and the enzyme concentrations were 0.3–0.5 μ M. Formation of *trans*-cinnamic acid and *p*-coumaric acid was followed spectrophotometrically by monitoring absorbance at 290 nm and 310 nm, respectively. Extinction coefficients of 9116 M⁻¹ cm⁻¹ (*trans*-cinnamic acid) and 11558 M⁻¹ cm⁻¹ (*p*-coumaric acid) were determined in the assay buffer and were used to calculate product concentrations according to Lambert Beers law. Kinetic parameters K_M and k_{cat} were obtained by fitting the Michaelis-Menten equation to a plot of the reaction rates versus substrate concentrations. For TcPAL, the hill equation was used instead.

Thermostability measurements. Thermostability measurements were performed on a Prometheus NT.Plex nanoDSF instrument (NanoTemper Technologies) in a 50 mM Tris-HCl, 125 mM NaCl buffer (pH 8.5). Protein unfolding was monitored by following the ratio of intrinsic protein fluorescence at 350 nm to 330 nm over time, increasing the temperature from 20 °C to 95 °C with 1 °C per minute. The melting temperature corresponds to the maximum of the first derivative of the 350/330 ratio. CD spectroscopy measurements were performed on a Chirascan-plus (Applied Photophysics) in a 10 mM Na₂HPO₄/NaH₂PO₄ buffer (pH 8.5) using $l = 0.1$ cm, 0.3 mL quartz cuvettes. A temperature ramp was set during which the temperature was increased with 0.5 °C per minute from 20 °C to 90 °C, and CD spectra were obtained between 200–260 nm at 1 °C intervals.

Incubation at 60 °C and activity measurements. For the prolonged 60 °C incubation, the enzymes were placed in a Thermomixer comfort (Eppendorf) at 60 °C for 30 and 60 minutes and were cooled to 37 °C before starting kinetic activity measurements. Enzymes that had been stored in 25 °C during the incubation were pre-heated to 37 °C for 15 minutes, as a reference for activity. Reactions were carried out in triplicates, in a total volume of 200 μ L at 37 °C with L-Phe as substrate (final concentration 4.5 mM). The final enzyme concentrations were between 0.4–0.5 μ M. Absorbance was measured spectrophotometrically at 290 nm.

Assessment of proteolytic stability. To assess susceptibility to pepsin cleavage, reactions of 150 μ L were set up with 67 μ g/ml pepsin (from pig gastric mucosa, Roche) dissolved in 10 mM HCl at pH 2.0 and 0.4 μ M (≈ 32 μ g/ml) of RgPAL or MEGA_A1, respectively. As a control, the enzymes were dissolved in 150 μ L 10 mM HCl without pepsin at the same concentration. Reactions were carried out in 96-well plates at 37 °C and samples were taken at $t = 0, 15, 35, 60, 90$ and 120 min. A part of the sample was then used for SDS-PAGE analysis and a part was used to measure activity for 5 minutes with 1 mM L-Tyr at 37 °C. To assess susceptibility to trypsin cleavage, reactions of 600 μ L were set up with 10 μ g/ml trypsin (Sequencing Grade Modified Trypsin, Promega) dissolved in 50 mM Tris-HCl, 125 mM NaCl buffer at pH 8.5 and 1 μ M (≈ 80 μ g/ml) of RgPAL or MEGA_A1, respectively. As a control, the enzymes were dissolved in 600 μ L buffer without trypsin at the same concentration. Reactions were placed at 37 °C at 600 rpm and samples were taken at $t = 12, 24, 36, 50, 80, 120, 160$ and 240 min. At each time point, the sample was taken and immediately mixed with 1 mg/ml trypsin inhibitor (Aprotinin from bovine lung, Sigma-Aldrich) to quench the reaction (a ratio of inhibitor to trypsin of ca. 6:1). A part of the sample was then used for SDS-PAGE analysis and a part was used to measure activity for 5 minutes with 700 μ M L-Tyr at 37 °C. To ensure that activity was completely lost in the samples with trypsin, a final sample was taken 20 hours after starting the reactions.

Activity measurements in serum. Enzymatic breakdown of L-Tyr in calf serum was measured at different concentrations of RgPAL and MEGA_A1. Calf serum was spiked with 800 μ M of L-Tyr. Enzyme solutions in a range of concentrations were added to the serum, resulting in a final reaction volume of 500 μ L and a final L-Tyr concentration of 775 μ M. The final enzyme concentrations were 2000, 1000, 500, 250, 125, 62.5, 31.25 and 0 nM. Reactions were carried out in a thermoshaker at 37 °C and 600 rpm. The vials with RgPAL were frozen at -80 °C after 90 minutes and the reactions with MEGA_A1 were frozen after 120 minutes. Samples were analyzed by RP-HPLC chromatography applying a sodium acetate/acetonitrile gradient. Pre-column derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) was employed and subsequent detection was done by UV absorbance at 254 nm.

Long-term stability assessment. For assessment of long-term stability of RgPAL, TcPAL and MEGA_A1, aliquots of 200 μ L were prepared with 1 mg/ml of enzyme in SEC buffer (50 mM Tris-HCl, 125 mM NaCl, pH 8.5). Samples were filtered using 0,22 μ m filters under sterile conditions and were incubated at either 25 °C or 37 °C. One aliquot of each enzyme was stored at -80 °C for reference. After 2 weeks, one sample per enzyme was withdrawn from 37 °C and stored at -80 °C. After 4 weeks additional samples were withdrawn from 25 °C and 37 °C and stored at -80 °C. All samples were thawed simultaneously and analyzed using SDS-PAGE and activity assays at 37 °C.

Data availability

Additional data is available in the Supplementary Information file.

Received: 19 July 2019; Accepted: 7 January 2020;

Published online: 28 January 2020

References

- Scott, R. C. The genetic tyrosinemas. *Am. J. Med. Genet.* **142C**, 121–126 (2006).
- Lindstedt, S., Holme, E., Lock, E. A., Hjalmarson, O. & Strandvik, B. Treatment of hereditary tyrosinaemia type I by inhibition of 4-hydroxyphenylpyruvate dioxygenase. *Lancet* **340**, 813–817 (1992).
- Bendadi, F. *et al.* Impaired cognitive functioning in patients with tyrosinemia type i receiving nitisinone. *J. Pediatr.* **164**, 398–401 (2014).
- Scriver, C. R. & Clow, C. L. Phenylketonuria: epitome of human biochemical genetics (first of two parts). *N. Engl. J. Med.* **303**, 1336–1342 (1980).
- Donlon, J., Starkissian, C., Levy, H. & Scriver, C. R. Hyperphenylalaninemia: phenylalanine hydroxylase deficiency in *The Online Metabolic and Molecular Bases of Inherited Disease* (eds. Scriver, C. R., Beaudet A. L., Sly, W. S. & Valle, D.) (McGraw-Hill Medical Publishing Division, New York, 2001).
- Williams, R., Mamotte, C. D. & Burnett, J. Phenylketonuria: an inborn error of phenylalanine metabolism. *Clin. Biochem. Rev.* **29**, 31–41 (2008).
- Wang, L. *et al.* Structure-based chemical modification strategy for enzyme replacement treatment of phenylketonuria. *Mol. Genet. Metab.* **86**, 134–140 (2005).
- Sarkissian, C. N. *et al.* A different approach to treatment of phenylketonuria: Phenylalanine degradation with recombinant phenylalanine ammonia lyase. *Proc. Natl. Acad. Sci.* **96**, 2339–2344 (1999).
- Isabella, V. M. *et al.* Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat. Biotechnol.* **36**, 857–867 (2018).
- Huisman G. W., Agard N. J., Mijts, B., Vroom, J. & Zhang, X. Inventors; Codexis, Inc., applicant. Engineering phenylalanine ammonia lyase polypeptides. United States patent US 0191050 A1. 2017 Jul 6.
- Kyndt, J. A., Meyer, T. E., Cusanovich, M. A. & Van Beeumen, J. J. Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Lett.* **512**, 240–244 (2002).
- Parmeggiani, F., Weise, N. J., Ahmed, S. T. & Turner, N. J. Synthetic and therapeutic applications of ammonia-lyases and aminomutases. *Chem. Rev.* **118**, 73–118 (2018).
- Poppe, L. Methylidene-imidazolone: A novel electrophile for substrate activation. *Curr. Opin. Chem. Biol.* **5**, 512–524 (2001).
- Cooke, H. A., Christianson, C. V. & Bruner, S. D. Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes. *Curr. Opin. Chem. Biol.* **13**, 460–468 (2009).
- Lovelock, S. L. & Turner, N. J. Bacterial *Anabaena variabilis* phenylalanine ammonia lyase: A biocatalyst with broad substrate specificity. *Bioorganic Med. Chem.* **22**, 5555–5557 (2014).
- Weise, N. J., Parmeggiani, F., Ahmed, S. T. & Turner, N. J. The bacterial ammonia lyase EncP: A tunable biocatalyst for the synthesis of unnatural amino acids. *J. Am. Chem. Soc.* **137**, 12977–12983 (2015).
- Garrait, G. *et al.* Gastrointestinal absorption and urinary excretion of trans-cinnamic and p-coumaric acids in rats. *J. Agric. Food Chem.* **54**, 2944–2950 (2006).
- Hoskins, J. A. & Gray, J. Phenylalanine Ammonia Lyase in the Management of Phenylketonuria: The Relationship Between Ingested Cinnamate and Urinary Hippurate in Humans. *J. Urol.* **11**, 986–989 (2017).
- Levy, H. L., Sarkissian, C. N. & Scriver, C. R. Phenylalanine ammonia lyase (PAL): From discovery to enzyme substitution therapy for phenylketonuria. *Mol. Genet. Metab.* **124**, 223–229 (2018).
- Rosler, J., Krekel, F., Amrhein, N. & Schmid, J. Maize Phenylalanine Ammonia-Lyase Has Tyrosine Ammonia-Lyase Activity. *Plant Physiol.* **113**, 175–179 (1997).
- Zhu, L. *et al.* Cloning, expression and characterization of phenylalanine ammonia-lyase from *Rhodotorula glutinis*. *Biotechnol. Lett.* **35**, 751–756 (2013).
- Vannelli, T., Xue, Z., Breinig, S., Qi, W. W. & Sariaslani, F. S. Functional expression in *Escherichia coli* of the tyrosine-inducible tyrosine ammonia-lyase enzyme from yeast *Trichosporon cutaneum* for production of p-hydroxycinnamic acid. *Enzyme Microb. Technol.* **41**, 413–422 (2007).
- Jendresen, C. B. *et al.* Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **81**, 4458–4476 (2015).
- Zakas, P. M. *et al.* Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat. Biotechnol.* **35**, 35–37 (2017).
- Gumulya, Y. & Gillam, E. M. J. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the retro approach to protein engineering. *Biochem. J.* **474**, 1–19 (2017).
- Risso, V. A. & Sanchez-Ruiz, J. M. Resurrected ancestral proteins as scaffolds for protein engineering in *Directed Enzyme Evolution: Advances and Applications* (ed. Alcalde, M.) 229–255 (Springer International Publishing AG, 2017).
- Hendrikse, N. M., Charpentier, G., Nordling, E. & Syrén, P.-O. Ancestral diterpene cyclases show increased thermostability and substrate acceptance. *FEBS J.* **285**, 4660–4673 (2018).
- Perica, T. *et al.* Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science* **346**, 1254346 (2014).
- Merkl, R. & Sterner, R. Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* **397**, 1–21 (2016).
- Aakre, C. D. *et al.* Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. *Cell* **163**, 594–606 (2015).
- Hochberg, G. K. A. *et al.* Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions. *Science* **359**, 930–935 (2018).
- Goldson-Barnaby, A. & Scaman, C. H. Purification and Characterization of Phenylalanine Ammonia Lyase from *Trichosporon cutaneum*. *Enzyme Res.* 1–6 (2013).
- Silva, D. N. *et al.* Genomic patterns of positive selection at the origin of rust fungi. *PLoS One* **10**, e0143959 (2015).

34. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
35. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
36. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
37. Martin, L., Schwarz, S. & Breitsprecher, D. Thermal unfolding Analyzing Thermal Unfolding of Proteins. *Nanotemper Technol. Appl. Note* 1–8 (2014).
38. Magnusson, A. O. *et al.* nanoDSF as screening tool for enzyme libraries and biotechnology development. *FEBS J.* **286**, 184–204 (2019).
39. Zhu, L., Zhou, L., Cui, W., Liu, Z. & Zhou, Z. Mechanism-based site-directed mutagenesis to shift the optimum pH of the phenylalanine ammonia-lyase from *Rhodotorula glutinis* JN-1. *Biotechnol. Rep.* **3**, 21–26 (2014).
40. Emiliani, G., Fondi, M., Fani, R. & Gribaldo, S. A horizontal gene transfer at the origin of phenylpropanoid metabolism: A key adaptation of plants to land. *Biol. Direct* **4**, 1–12 (2009).
41. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
42. Straub, K. *et al.* Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranyl glyceryl phosphate synthase. *Biol. Chem.* **400**, 367–381 (2018).
43. Poppe, L. & Rétey, J. Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine. *Angew. Chem. - Int. Ed.* **44**, 3668–3688 (2005).
44. Pilbák, S., Farkas, Ö. & Poppe, L. Mechanism of the tyrosine ammonia lyase reaction-tandem nucleophilic and electrophilic enhancement by a proton transfer. *Chem. - A Eur. J.* **18**, 7793–7802 (2012).
45. Miller, S. R. An appraisal of the enzyme stability-activity trade-off. *Evolution* **71**, 1876–1887 (2017).
46. Rosenberg, A. S. Effects of protein aggregates: An immunologic perspective. *AAPS J.* **8**, 501–507 (2006).
47. Moffitt, M. C. *et al.* Discovery of two cyanobacterial phenylalanine ammonia lyases: Kinetic and structural characterization. *Biochemistry* **46**, 1004–1012 (2007).
48. Wang, L. *et al.* Structural and Biochemical Characterization of the Therapeutic *Anabaena variabilis* Phenylalanine Ammonia Lyase. *J. Mol. Biol.* **380**, 623–635 (2008).
49. Perica, T., Chothia, C. & Teichmann, S. A. Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl. Acad. Sci.* **109**, 8127–8132 (2012).
50. Goyal, V. D., Sullivan, B. J. & Magliery, T. J. Phylogenetic Spread of Sequence Data Affects Fitness of Consensus Enzymes: Insights from Triosephosphate Isomerase. *Proteins Struct. Funct. Bioinforma.* **24**, 1–10 (2019).
51. Vieille, C. & Zeikus, G. J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43 (2001).
52. Koudelakova, T. *et al.* Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chemie - Int. Ed.* **52**, 1959–1963 (2013).
53. Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260 (1994).
54. Yu, H., Yan, Y., Zhang, C. & Dalby, P. A. Two strategies to engineer flexible loops for improved enzyme thermostability. *Sci. Rep.* **7**(41212), 1–1 (2017).
55. Gilbert, H. J. & Jack, G. W. The effect of proteinases on phenylalanine ammonia-lyase from the yeast *Rhodotorula glutinis*. *Biochem. J.* **199**, 715–723 (1981).
56. Kim, W. *et al.* Trends in enzyme therapy for phenylketonuria. *Mol. Ther.* **10**, 220–224 (2004).
57. Huang L. L., Xue Z., McCluskey M. P., inventors; E. I. du Pont de Nemours and Company, assignee. Method of production of parahydroxycinnamic acid using a thermostable TAL enzyme. United States patent US 7,572,612 B2. 2009 Aug 11.
58. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
59. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
60. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
61. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
62. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32** (2015).
63. Suchard, M. A. & Redelings, B. D. BALI-Phy: Simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**, 2047–2048 (2006).
64. Venselaar, H. *et al.* Homology modelling and spectroscopy, a never-ending love story. *Eur. Biophys. J.* **39**, 551–563 (2010).

Acknowledgements

We acknowledge the financial support of the Swedish Foundation for Strategic Research (SSF), grant number ID16-0036. We thank Robert Svensson for assistance with the long-term stability study and we thank Karin Andersson for her input in the project. We also thank Jan Kinnander for analysis of the amino acid concentration in serum. Open access funding provided by Royal Institute of Technology.

Author contributions

E.N., S.S.G. and P.O.S. designed part of the study and supervised the work. N.M.H. designed part of the study, performed part of the experiments and wrote the major part of the manuscript. A.H.L. performed part of the experiments. S.K. assisted with the C.D. experiments and performed analysis and interpretation of data.

Competing interests

Stefan Svensson Gelius, Sergei Kuprin and Erik Nordling are employed by Swedish Orphan Biovitrum AB. Natalie Hendrikse is funded by the Swedish Foundation for Strategic Research and jointly employed by the KTH Royal Institute of Technology and Swedish Orphan Biovitrum AB. Per-Olof Syrén and Albin Holmberg Larsson declare no conflict of interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-57913-y>.

Correspondence and requests for materials should be addressed to P.-O.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020