ChemBioChem

Research Article
doi.org/10.1002/cbic.202401028

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

Expanding the Enzymatic Toolbox for Carboligation: Increasing the Diversity of the 'Split' Transketolase Sequence Space

Alessia Tonoli,[a] Silvia Anselmi,[b] John M. Ward,[a] Helen C. Hailes,[b] and Jack W. E. Jeffries*[a]

Transketolases (TKs) are thiamine diphosphate (ThDP)-dependent enzymes that catalyze the transfer of two-carbon units in a stereoselective manner, making them valuable biocatalysts for sustainable processes. Most known TKs are about 650 amino acids long; however, a second type found in Archaea and many Bacteria consists of two proteins, each of about 300 amino acids. Exploring the unique features and differences of split TKs may help in assessing their potential use in biocatalysis and for uncovering new reactivities. Additionally, it could provide valuable information on how their structure relates to their function, especially compared to full-length TKs. In this study, we significantly expanded the known repertoire of split TKs approximately 14-fold to the best of our knowledge, by identifying and providing accessions of nearly 500 putative split-TK subunit pairs. Moreover, we doubled the number of experimentally produced and tested split TKs by cloning, purifying, and testing ten candidates retrieved from genomes and in-house metagenomes. Interestingly, pQR2809 and pQR2812, derived from hyperthermophilic organisms, showed enhanced thermostability compared to other TK examples in the literature, maintaining partial activity after heating at 90 °C or 100 °C for 1 hour, respectively.

Introduction

The enzyme transketolase (TK) (EC 2.2.1.1) plays a key role in the pentose phosphate pathway, catalyzing the transfer of a two-carbon ketol unit from a ketose donor to an aldose acceptor, leading to the elongation of the carbon backbone in a stereoselective fashion.[1] Different TKs, including those from *Escherichia coli* (EcoTK) and *Geobacillus stearothermophilus* (GstTK), have undergone extensive characterization studies[2] and mutagenesis efforts to enhance their stability,[3] the acceptance of various unnatural substrates,[4] and to improve or reverse their stereospecificity,[5] ultimately optimizing the production of desired products. Among the diverse TKs reported in literature, there are some thermostable enzymes which have promise for industrial biocatalysis applications.[2c,6] Notably, one of these is the TK from the hyperthermophilic bacterium *Carboxydothermus hydrogenoformans* (ChyTK) which is able to endure extreme temperatures and high percentages of organic solvents.[6b] However, this enzyme is markedly distinct from the previously mentioned TKs, which are homodimeric enzymes with two active sites at the interface of the dimer. In the ChyTK the full-length TK monomer is split across separate open reading frames into two distinct proteins of around 300 amino acids, instead of a single subunit of around 650 amino acids. The two active sites in this enzyme are created at the interface of the heterotetramer. It was therefore defined as a 'split-gene' transketolase (split TK), and its crystallographic structure was solved (PDB ID: 6yak) (Figure 1).[6b] It was hypothesized that the different architecture of split TKs could

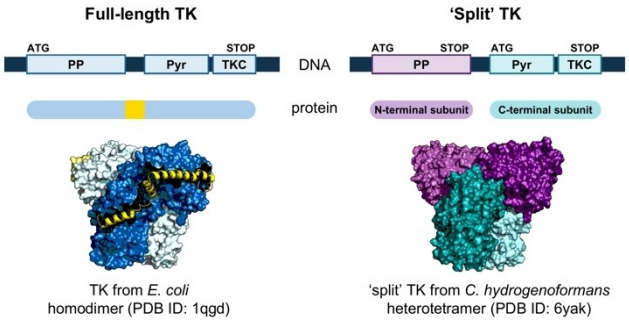


Figure 1. Comparison of domains and subunit organization of full-length TKs versus split TKs, with reference structures. Full-length TKs (e.g. EcoTK, PDB ID: 1qgd[2b]) are translated as single subunits of around 650 amino acids, which then form homodimers. 'Split' TKs (e.g. ChyTK, PDB ID: 6yak[6b]) are translated as two separate subunits, which then form heterotetramers. The peptide linking the PP domain to the Pyr domain in full-length TKs (in yellow) is absent in split TKs. ATG/STOP = start/stop codon for transcription of messenger RNA from the coding sequence. PP = pyrophosphate binding domain; Pyr = pyrimidine binding domain; TKC = transketolase C-terminal domain; N/C-terminal = amino/carboxy terminal subunits of split TK. The structures of the proteins were rendered using PyMOL Molecular Graphics System, Schrödinger, LLC.

[a] *A. Tonoli, J. M. Ward, J. W. E. Jeffries*
Department of Biochemical Engineering, University College London, Bernard Katz Building, Gower Street, London W1CE 6BT, United Kingdom
E-mail: jack.jeffries.12@ucl.ac.uk

[b] *S. Anselmi, H. C. Hailes*
Department of Chemistry, University College London
Christopher Ingold Building, 20 Gordon Street, London WC1H 0AJ, United Kingdom

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202401028

allow greater flexibility, for the potential accommodation of larger or bulkier substrates compared to full-length TKs.[6b]

Other TKs with the split architecture have been reported in the literature, although most often they are not labelled as *split*. A dozen of these were produced and experimentally tested.[7] Two examples are the D-apulose 4-phosphate TK AptAB, which has also been assigned an EC number (EC 2.2.1.13), and the 3-oxo-isoapionate TK OitAB. These enzymes have been proposed to be involved in the catabolic pathways of D-apiose and D-apionate, respectively, using either D-apulose 4-phosphate (in the case of AptAB) or 3-oxo-isoapionate (in the case of OitAB) as donors, and D-glyceraldehyde 3-phosphate as acceptor.[7b] A further case of a split TK homologue, FtxE/F, was found to participate in the reconstituted biosynthetic pathway for the production of phosphonothrixin (PTX), a molecule with herbicidal properties.[7h,i]

More recently, a TK involved in sulfur-recycling pathways has been described to participate in the cleavage of sulfoquinovose, a component of green plant sulfolipids, using this sugar and its degradation product 4-deoxy-4-sulfoerythrulose as ketol donors, with glyceraldehyde 3-phosphate as acceptor.[7f,g]

Metagenome mining has expanded the repertoire of enzymes available, revealing an enzymatic landscape previously unexplored.[8] Our group some years ago reported the retrieval of five TKs from a metagenome derived from the oral cavity.[7a] Among the two TKs characterized, one contained a stop codon within its open reading frame, constituting another example of a split TK.

Here, we aimed to broaden the panel of split TKs with potential applicability in biocatalysis by exploring this class of TKs and their diversity. Accordingly, we report a novel array of split TKs identified through both genome and metagenome mining. Ten of these enzymes were successfully produced and tested for their TK-activity and thermostability, in comparison to TKs characterized in literature.

Results and Discussion

Novel Putative Split TKs from Genomes and Metagenomes

To expand the assortment of split TKs available for investigation, both genomic and metagenomic repositories were explored. Five in-house metagenome databases were investigated for the presence of putative split TKs. Three of these derive from Peruvian salt mines located in the Amazon or in the highlands (Pilluana, 6Maras and Maras3), while the other two samples (MV16 and MV17) originate from the soil of the dry Miers Valley in Antarctica.[9] Salt mines are valuable sources for the discovery of novel enzymes derived from organisms that have evolved to withstand extreme conditions of salinity and unique environmental stresses, therefore potentially able to tolerate harsh conditions often needed in industrial processes.[10] On the other hand, Antarctica dry valleys are environments with unique microorganisms,[11] and could represent an important reservoir for the discovery of novel enzymes. Recently, a DERA (deoxy ribose-phosphate aldolase) derived from one of these metagenomes was reported for its biocatalytic promiscuity.[12]

Our group has previously used sequence-based metagenomic approaches utilizing Pfam identifiers[13] to extract various enzymes, such as transaminases,[14] ene-reductases,[15] and carbonyl reductases.[16] Building on this methodology, in the current study, we initially scanned datasets of Pfam-annotated open reading frames (ORFs) to detect TK-typical domains through their identifiers. To gather enzymes with a split gene and differentiate them from full-length ones, sequences were picked in which the thiamine diphosphate binding domain (PF00456) was located on an ORF distinct from the one containing the other two domains – pyrimidine binding domain (PF02779) and transketolase C-terminal domain (PF02780). Starting from more than 3.6 million contigs including >4.2 million ORFs across the 5 metagenomes analyzed, only 157 contigs contained all three TK domains (Table S1). Of these, 39 displayed the split pattern, of which 17 constituted complete genes (with a start and a stop codon). These were all selected for expression. This data shows that already in this limited dataset, a quarter of the putative enzymes containing these three domains are split, suggesting that this architecture is not uncommon. The majority of the retrieved contigs contained putative full-length transketolases (70 complete genes) (Figure 2A).

Further examples of putative split TKs were also retrieved from the annotated genomes of various Bacteria and Archaea. Focusing on various mesophilic and thermophilic organisms of interest, 10 enzymes derived from online databases were chosen for further analysis.

Altogether, a total of 27 split TKs were selected for expression in *Escherichia coli*. A plasmid ID (pQR number) was assigned to each of these proteins. Details of these enzymes, along with their genomic origin or their closest homologue, are provided in Tables S3–S6. The existence of some other of these putative split TKs was reported in literature (details in Table S4). This panel included the previously reported split ChyTK (pQR2806), characterized by James and colleagues,[6b] as well as the split TK from *Saccharolobus solfataricus* (pQR2814), which had already been expressed and tested for activity in unpublished work from our group.[17] Aside from these examples, to the best of our knowledge, the other split TKs described herein have not been produced or investigated before.

One of the sequences (pQR2807) when synthesized inadvertently constituted a chimera between two subspecies of *Fusobacterium nucleatum* (subspecies *nucleatum* and *polymorphum*). Compared to the wild-type enzyme from *F. nucleatum* subsp. *nucleatum*, this chimera differs in the C-terminal subunit by five amino acids, which, based on sequence alignment and the literature, are likely not catalytically relevant (Figure S1). However, the correct plasmid with the native sequence was also produced (pQR2834) and compared to its chimeric version.

The percent identity matrix of split TKs selected for production and reference full-length ones (Figures S2-S4) indicates that these enzymes vary widely in their similarity levels. The degree of identity among all selected split TKs ranged from 27 % to 85 %. Some split TKs formed clusters with a high mutual

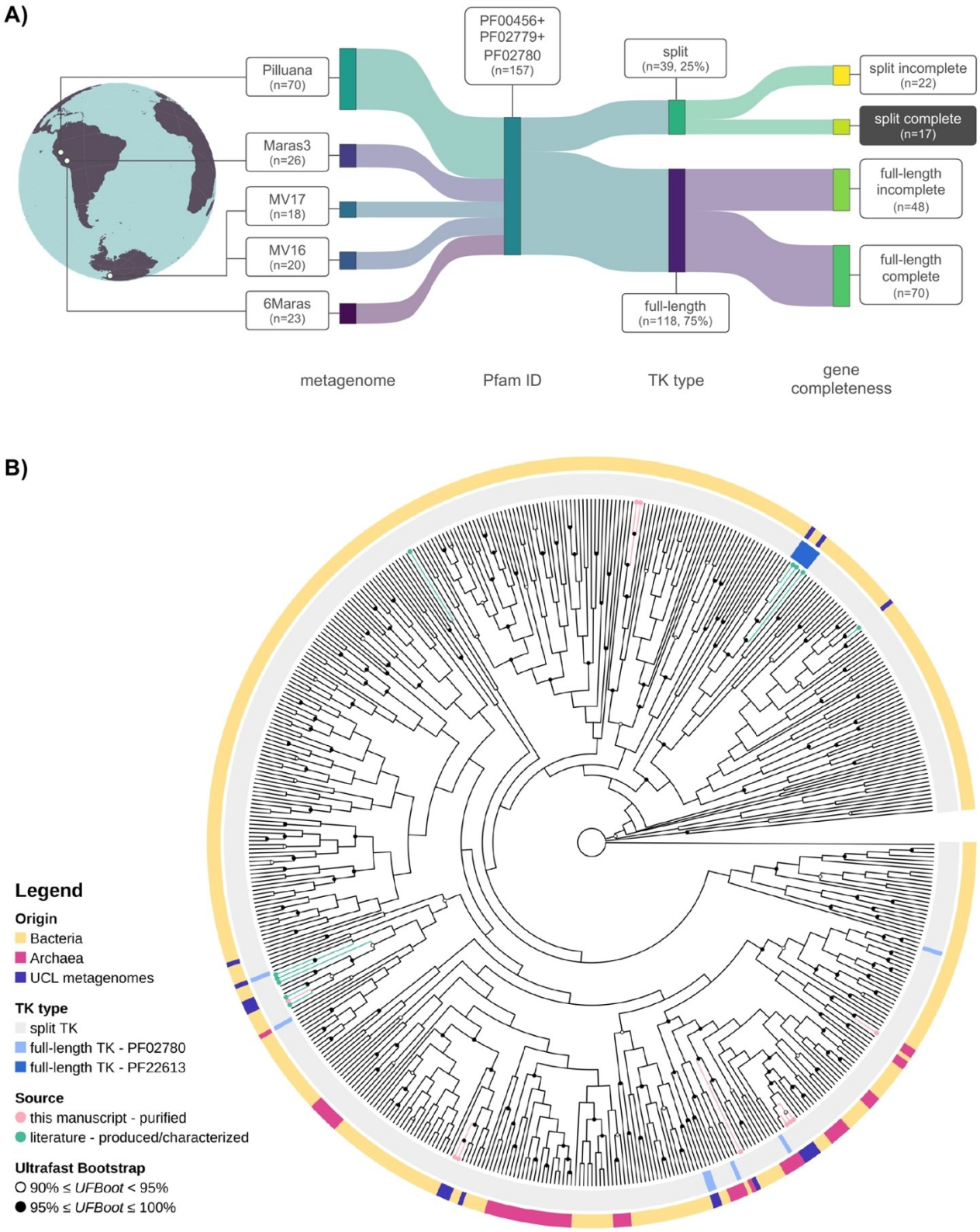


Figure 2. A) Sankey diagram illustrating the workflow for the identification of putative split TKs from in-house metagenome datasets. Firstly, all the sequences containing all three TK-typical domains (according to Pfam 35.0 or previous versions) were retrieved (157 sequences). Secondly, these were filtered for the presence of the target domains on a single coding sequence (full-length TKs, 118) or on two separated sequences ('split' TKs, 39). Finally, only the complete sequences were considered (17 complete split enzymes, 70 complete full-length enzymes). B) Unrooted phylogenetic tree including retrieved split TKs. The putative split TKs successfully purified are highlighted (pink terminal nodes) (except for pQR2815, not included in this analysis), and derive from genomes of Bacteria (yellow), Archaea (magenta), or from in-house metagenomes (purple). Sequences of full-length TKs (12), including some sourced from the in-house metagenomes, were also included. Black or white internal nodes indicate an ultrafast bootstrap value higher than 95%, or between 90% and 95%, respectively.

degree of identity, of over 50%, suggesting that they are closely related.

However, when compared to full-length TKs, these enzymes show lower identity - less than 32% with EcoTK and less than 40% with all full-length TKs used as a reference. This highlights both the variability within the group of split-gene TKs alone and their divergence from full-length TKs.

From the original set of split TKs, a sequence-based approach was employed to expand the list and further explore the diversity and phylogeny of this class of enzymes. Overall, this research led to the retrieval of a total of 505 split TK pairs – including the enzymes from the literature – of which, to the best of our knowledge, 491 putative candidates have not been reported before. Moreover, this investigation also resulted in the retrieval of some putative full-length TKs from Archaea (Table S8).

These split TKs – all checked for the presence of the three TK-typical domains – were gathered in a phylogenetic tree, together with some examples of full-length TKs, including some derived from the UCL proprietary metagenomes (Figure 2B). The widespread distribution across the tree of the split TKs selected for further investigation underscores our strategy of retrieving diverse TKs for characterization, aiming to provide a more comprehensive understanding of this heterogenous group of enzymes. Full-length TKs do not form a single clade but are instead distributed into smaller clusters throughout the phylogenetic tree, interspersed with split TKs. These smaller groupings are consistent with recent updates in the Pfam database, which now distinguishes the PF22613 domain (Transketolase-like C-terminal domain) from the previously unified PF02780 (Transketolase C-terminal domain). In earlier Pfam versions, specifically version 35.0 and earlier, both full-length and split TKs were categorized under the PF02780 domain. However, since version 36.0, many full-length TKs have been reclassified under PF22613, while most split TKs continue to be annotated under PF02780. It has been previously proposed that full-length TKs may derive from a heterotetrameric ancestor, with a domain organization as that of split TKs.[18] Our analysis suggests the possibility that different full-length TKs might have originated from multiple independent fusion events of split TK genes.

Expression Strategy, Protein Synthesis and Purification

The strategy employed for the expression of putative split TKs consisted of a synthetic polycistron. Specifically, a stop codon was placed between the two TK subunits, downstream the first coding sequence (CDS), followed by a second ribosome binding site (RBS) and a start codon at the beginning of the second CDS. In two cases, the start codon of the N-terminal subunit was originally GTG (valine), but it was replaced with ATG (methionine) for expression in *E. coli*. The CDSs of the two subunits, with this synthetic sequence in-between, were codon optimized for expression in *E. coli* and then ordered from GenScript Biotech (UK), cloned within a commercial pET-29b(+) vector (Figure S5). This system allowed for the transcription of a single polycistronic messenger RNA from the T7 promoter, which would then be translated from the two RBSs into two distinct polypeptides. In the literature, another split TK has been produced exploiting the insertion of a second synthetic RBS between the two ORFs.[7f] Although the polycistronic expression system cannot guarantee the production of both subunits in equal stoichiometry, it mimics the native organization of the subunits of split TKs, which frequently occur in a single operon.

After transformation of *E. coli* BL21(DE3) with the plasmids, different expression conditions (media and expression temperatures) were tested, but only 11 enzymes out of 27 (41%) were successfully overexpressed in soluble form. The split TK from *C. hydrogenoformans* could not be produced at high levels with the polycistronic system. In many cases, there were solubility issues concerning one or both proteins (37%). In other instances, one or both subunits showed low expression (11%) or no expression at all (11%) (Table S9).

The subunit downstream in the expression construct was kept in frame with the His_6-tag already present in the vector, to allow for purification of the enzyme through immobilized metal affinity chromatography (IMAC). All overexpressed, soluble split TKs were successfully purified, with few impurities, except for pQR2814 (derived from *S. solfataricus*) (Figure 3).

In constructing the plasmids, the native order of subunits (N-terminal subunit upstream, C-terminal subunit downstream) was preserved. In the case of the split TK from *Methanocaldococcus jannaschii*, the plasmids with both configurations were designed and tested (subunits in native order in pQR2816, in reversed order in pQR2812). In both cases, the enzyme was overexpressed.

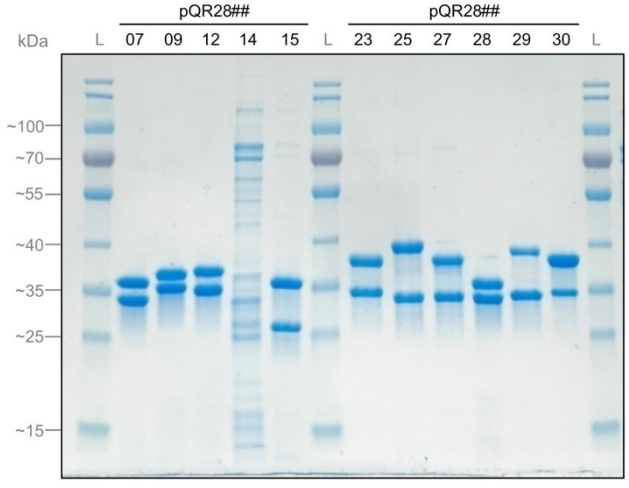


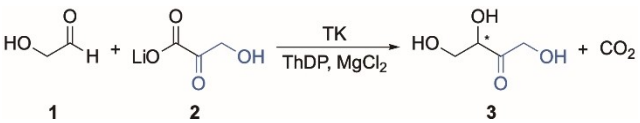
Figure 3. 12% SDS-PAGE analysis of elution fractions from the IMAC purification of split TKs. Protein purification was successful for all expressed split TKs, except for pQR2814, derived from *S. solfataricus*. pQR28##: plasmid ID/split TK. L: PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific).

Activity at Different Temperatures

The ten purified split TKs were tested for activity in the TK synthetic reaction with lithium hydroxypyruvate **2** as ketol donor, and glycolaldehyde **1** as the acceptor (Scheme 1). The use of **2** in the TK reaction has been widely employed in the literature in order to drive the reaction to completion with the release of carbon dioxide, compared to the reversibility of typical TK native reactions.[19] Using **2** to test split TKs also served as an early assessment of their synthetic potential. Additionally, the synthesis of erythrulose **3** can be monitored using HPLC analysis and detection at 210 nm (or using an RI detector), as performed in previous studies,[4a,6c] and it is commercially available. The reaction was tested at 3 different temperatures – 25 °C, 30 °C, and 50 °C – providing an initial assessment of the thermostability of these enzymes. The reactions were performed for 1 hour to identify active enzymes, while also limiting the possibility of reagent degradation. Enzymes were also assayed at 25 °C following a 20-hour incubation to allow detection in case of low activity (Figure S6).

From this analysis, purified native pQR2834 and chimeric pQR2807 displayed comparable activities and gave the highest yields overall, calculated by HPLC against product standards, ranging from 47% to 58% after 1 hour at each temperature tested. pQR2829 achieved 47% yield at 25 °C after 1 hour, but exhibited lower yields at higher temperatures, suggesting poor thermostability. A similar pattern was observed with pQR2830, but with lower overall yields (7% yield after 1 hour at 30 °C). In contrast, pQR2809, pQR2812 and pQR2823 displayed higher activity at 50 °C, compared to 25 °C and 30 °C. pQR2812 and pQR2816, which differ only in the position of the His-tag, showed similar yields to each other (Figure S9). pQR2825, pQR2827 and pQR2828 showed only traces of product or no activity with these substrates under the conditions tested, even after 20 hours at 25 °C. pQR2815 displayed minimal activity after 1 hour of reaction but achieved 9% yield after 20 hours at 25 °C (Figure 4 and Figure S6).

Among the tested enzymes, pQR2834, along with its chimeric version pQR2807, demonstrated the highest yields after 1 hour of reaction and showed potential thermostability. Consequently, pQR2834 was selected for further investigation to determine its optimal temperature and pH conditions. This split TK, originated from the mesophilic bacterium *F. nucleatum*, displayed higher initial rates in the range of 37–50 °C and at pH 7.5–8.5 (Figure S7). After having identified its pH and temperature optima, its apparent kinetic profile was investigated for the substrate HPA **2** at 37 °C in 50 mM Tris-HCl buffer pH 7.8 (Figure S8). The kinetic data were fitted to the Hill



Scheme 1. TK-catalyzed reaction with lithium hydroxypyruvate (HPA, **2**) as ketol donor, and glycolaldehyde (GA, **1**) as acceptor, for the production of erythrulose (ERY, **3**).

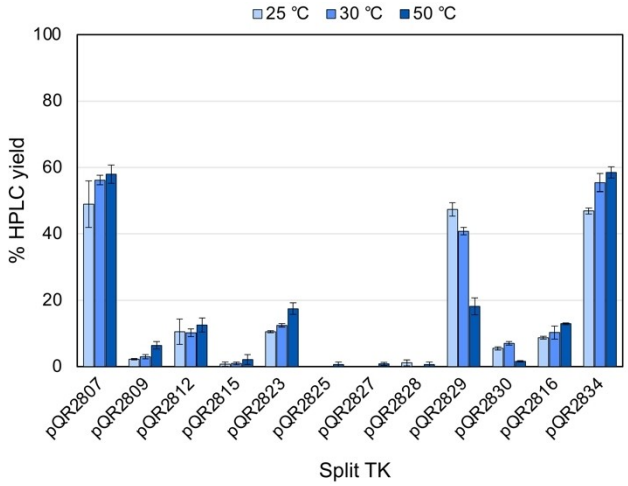


Figure 4. HPLC yields of split TK-catalyzed reaction with lithium hydroxypyruvate (HPA, **2**) and glycolaldehyde (GA, **1**) at different temperatures. Reactions were run at 25 °C, 30 °C and 50 °C for 1 h, using 0.05 mg/mL of purified split TKs. Reactions were run at least in triplicate.

function, showing positive cooperativity ($n = 1.4 \pm 0.1$), an affinity of 11 ± 1 mM, and a catalytic efficiency of 6.9 ± 0.4 s^{-1} mM^{-1}. No kinetic data for other split TKs related to donor **2** is currently available in the literature. In comparison to other full-length TKs, pQR2834 showed lower affinity than EcoTK and SceTK, but similar catalytic efficiency (Table 1).[20]

Thermostability of Split TKs

TKs that did not display any loss of activity when tested at 50 °C compared to 25 °C or 30 °C (i.e. pQR2807, pQR2809, pQR2812, pQR2823 and pQR2834) were also incubated for 1 hour at different temperatures in the range of 50–100 °C, rapidly cooled on ice, and re-tested to obtain an indication of their thermostability (Figure 5A).

While pQR2823 lost all its activity after incubation at 50 °C, both pQR2807 and pQR2834 were able to retain some up to 60 °C, loosing completely their activity after incubation at 70 °C. Interestingly, the HPLC yields after pre-incubation at 60 °C suggest that the chimeric version (pQR2807) is slightly more thermostable compared to the native protein (pQR2834), which

Table 1. Apparent kinetic parameters towards lithium hydroxypyruvate (HPA, **2**) of various TKs (split or full-length).

TK	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})	Reference
pQR2834[a]	11[d]	6.9	this manuscript
EcoTK[b]	5.5	6.0	[20]
SceTK[b,c]	3.4	8.3	[20]
TmaTK[b]	54	0.014	[6c]

[a] Split TK. [b] Full-length TKs. [c] TK from *Saccharomyces cerevisiae*. [d] Equivalent parameter in the Hill equation.

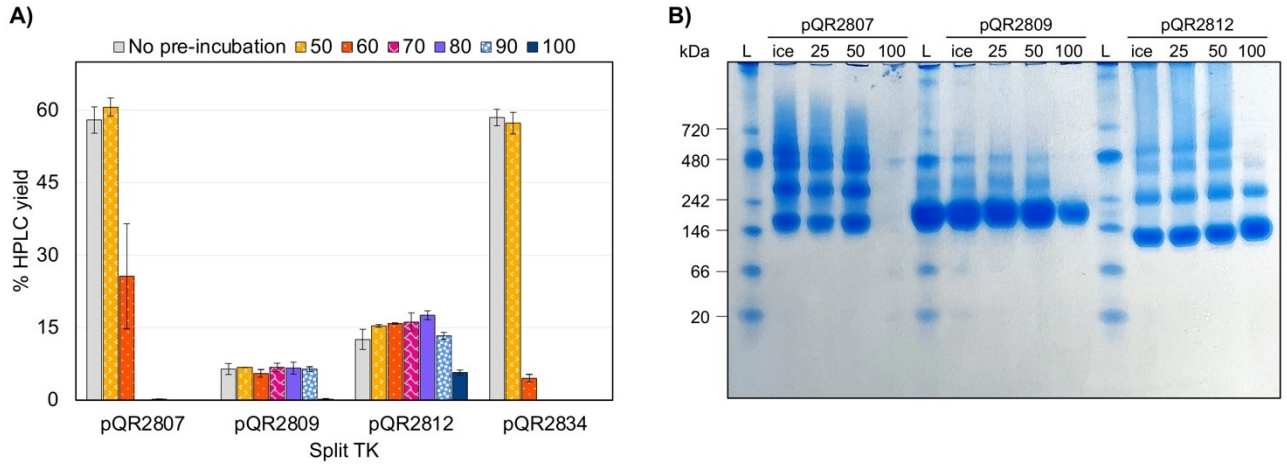


Figure 5. Thermostability of split TKs. A) Split TKs were tested in the reaction with GA **1** and HPA **2** after pre-incubation for 1 hour at different temperatures in the range between 50 °C and 100 °C. B) 4–20 % native PAGE of thermostable split TKs after pre-incubation at different temperatures. L: NativeMark™ Unstained Protein Standard (Invitrogen).

differs by only 5 amino acids in the C-terminal subunit. pQR2809 and pQR2812, both derived from thermophilic organisms, demonstrated marked traits of stability at higher temperatures. pQR2809, derived from the hyperthermophilic bacterium *Thermotoga maritima*, which optimally grows at 80 °C,[21] retained almost complete activity until incubation at 90 °C. The full-length TK from the same hyperthermophilic bacterium (TmaTK) showed an optimum temperature higher than 90 °C.[6c] pQR2812, from *M. jannaschii*, a hyperthermophilic archaea with an optimum growth temperature of 85 °C,[22] was able to retain some activity even after incubation at 100 °C.

Nevertheless, these last two split TKs exhibited low HPLC yields compared to other examples, which may suggest a trade-off between activity and stability, or possibly the employment of suboptimal substrates and/or reaction conditions. These data suggest that both enzymes can withstand higher temperatures compared to GstTK and ChyTK, which were reported to lose activity after incubation at 85 °C and 90 °C, respectively.[2c,6b] Moreover, pQR2812 outperformed the TK from *Thermus thermophilus* (TthTK), which completely loses its activity after incubation at 100 °C for 10 minutes.[6d] pQR2829 and pQR2830 were also tested for their thermal stability, but did not display any residual activity after incubation at 60 °C for 1 hour.

Split TKs pQR2807, pQR2809, and pQR2812 were also subjected to native PAGE analysis, after incubation at 25 °C, 50 °C and 100 °C for 30 minutes (Figure 5B). Native PAGE allows the observation of protein complexes and oligomeric states, and in this case, to detect any changes occurring after incubation at progressively higher temperatures. This analysis suggested that the minimum oligomeric state in solution for all three split TKs is a tetramer (protein bands at around 146 kDa, according to the native protein ladder), similarly to the reconstituted split enzyme ChyTK. However, multiple protein bands at higher apparent molecular weights suggests the presence of higher oligomeric states, such as octamers and hexadecamers, which tend to decrease in favor of the lowest oligomeric state as the temperature increases.

Reflecting the previous thermostability analysis, pQR2807 was completely denaturated after 1 hour at 100 °C and no protein band could be observed; after the same thermal treatment, pQR2809 was present only in the smaller oligomeric state (tetramer), in a reduced quantity, while the temperature increase caused the dissociation of higher molecular weight forms; for pQR2812, instead, the higher oligomeric states were absent, and there was a higher proportion of lower oligomeric states (tetramers and octamers) compared to incubations at lower temperatures.

Conclusions

In this study, we identified 491 novel putative split TKs from a combination of genomic and metagenomic sources, using both approaches to exploit natural diversity for enzyme discovery. This represents approximately a 14-fold increase in the number of retrieved putative split TKs, including accessions of both N- and C-terminal subunits. Additionally, we sourced examples of split TKs from literature, which are not straightforward to identify, due to inconsistent naming conventions. The phylogenetic analysis revealed the heterogeneity within this group of TKs, and this study suggests that TKs with the split domain architecture are more prevalent and varied than previously recognized. We successfully purified and tested ten split TKs for their biocatalytic potential, doubling the existing number of produced and tested split TKs, thus expanding our understanding of this underexplored enzyme group. Notably, two enzymes which originate from hyperthermophilic organisms – pQR2809 from *T. maritima* and pQR2812 from *M. jannaschii* – exhibited enhanced high-temperature tolerance compared to other TK examples from the literature. These split TKs potentially offer additional resources for industrial applications, where enzyme stability is paramount.

Experimental Section

Chemicals

All chemicals were purchased from Sigma-Aldrich, unless otherwise stated. ThDP stocks were neutralized with NaOH.

Mining of Putative Split TKs from Genomes and Metagenomes for Expression

Mining from proprietary metagenomes was carried out by retrieving the nodes (contigs) that included all three TK-typical domains from datafiles containing the ORFs already annotated according to the Pfam notation of protein families. According to Pfam 35.0 and previous versions, the nodes were then filtered to obtain the ones in which PF00456 (Transketolase_N) and PF02779 + PF02780 (Transket_pyr and Transketolase_C, respectively) were found on distinct CDSs.

Mining from target genomic repositories was performed by text-based search for the query "transketolase" into NCBI databases[23] of bacterial and archaeal genomes and assemblies, or UniProt/GTDB databases,[24] and selecting for TKs with short sequences, either DNA (~900 nucleotides) or protein (~300 amino acids). The retrieved putative sequences were verified with InterProScan tool for the presence of the peculiar TK Pfam domains.[25]

Dataset Expansion and Phylogenetic Analysis

A dataset of putative split TKs was built based on the presence of the TK-typical Pfam IDs located on two different ORFs. Initially, the N-terminal subunits and the corresponding C-terminal subunits of a group of split TKs (the ones selected for expression and the ones found in the literature) were separately aligned using Clustal Omega with default parameters.[26] The resulting alignments were used as queries for a search using HMMER (version 3.3)[27] against the nr90 database, with an E-value cutoff of 1e–10 and a maximum of 1,000 hits. After a first round of sequence clustering at 75% identity and 80% coverage with MMseqs2,[28] the reduced dataset was refined by: (i) selecting sequences within length ranges appropriate for split TKs (N-terminal subunits: 200–350 amino acids; C-terminal subunits: 225–375 amino acids); (ii) scanning the sequences using pfam_scan[29] and Pfam version 35.0, and removing those lacking the target domains, or containing additional non-TK domains. For sequences where only one subunit was initially retrieved, the neighbouring ORFs on the CDS assembly were inspected and the corresponding N-terminal or C-terminal subunit retrieved and identified using custom Python scripts (SI) and pfam_scan. Subsequently, a second round of dataset clustering at 80% identity and 80% coverage was performed using MMseqs2, after concatenation of the protein sequence pairs. Following de-concatenation and re-integration of the proteins of interest in the dataset (split TKs selected for expression and TKs from literature), the sequences were re-analyzed using pfam_scan (Pfam 35.0) to locate domain boundaries and extract the corresponding sequence portions. The extracted domains were separately aligned using MAFFT (version 7) with default parameters.[30] Clustal Omega, HMMER, MMseqs2 and MAFFT were accessed via the MPI Bioinformatics Toolkit.[31] The aligned domain sequences were visually inspected and manually curated with Jalview (version 2.11.4.0),[32] then trimmed using trimAl[33] ('automated1' setting), and finally concatenated. Poor quality or highly divergent sequences were excluded from the dataset before downstream analysis.

For phylogenetic reconstruction, IQ-TREE (version 2.3.6) was used to generate a maximum likelihood (ML) tree.[34] The following parameters were applied: '-m MFP' for automatic model selection, '-B 1000' for ultrafast bootstrap with 1000 replicates, and '-alrt 1000 -bnni' for further support. The best-fit substitution model, LG + I + R10, was selected based on the Bayesian Information Criterion (BIC). The resulting phylogenetic tree was visualized and formatted using TreeViewer (version 2.2.0).[35]

Cloning, Expression and Purification of Putative Split TKs

All the coding sequences were first codon optimized for production in *E. coli*. A synthetic RBS was interposed between the coding sequences corresponding to the two subunits, in order to create a polycistronic system. These sequences were purchased from GenScript (UK) as synthetic DNA cloned into a pET29b(+) plasmid between NdeI and XhoI restriction sites. Plasmid pQR2834 was constructed via Gibson assembly by amplifying the shared portion from pQR2807, and the differing segment from a synthetic gene. Further details are provided in the Supporting Information.

All plasmids were transformed into *E. coli* BL21(DE3) cells (Invitrogen) for gene expression according to the manufacturer's protocol.

The standard expression of split TKs was started with the inoculation from the bacterial glycerol stock of a 1–10 mL-overnight culture containing LB (Luria-Bertani Broth) and the antibiotic kanamycin (Bio Basic) at final concentration of 50 μg/mL from a 1000X stock. The overnight culture was diluted 1 : 100 to set up a main culture in varying final volumes depending on the purpose: for screening of expression conditions, the culture was set up either in 96-well deep-well plates (500 μL culture) or in 50 mL-tubes (10 mL of culture); for large-scale cultures, either 1 L- or 2 L-flasks were used (culture volume equal to $^1/_5$ of maximum capacity). Expression conditions (temperature and medium) used for each split TK are detailed in Table S9. The culture was grown in a shaking incubator (New Brunswick Innova® 43) at 37 °C, 250 rpm, while in an Eppendorf ThermoMixer with ThermoTop at 1000 rpm when using 96-well plates. When the optical density at 600 nm (OD_{600}) reached $OD_{600} \geq 0.5–0.6$ when using LB – or $OD_{600} \geq 0.8–1.0$ when using TB (Terrific Broth, Merck Millipore) – IPTG (isopropyl ß-d-1-thiogalactopyranoside) was added to a final concentration of 1 mM from a 1000X stock to induce expression. The main culture was then placed in the shaking incubator at desired temperature, 250 rpm, overnight (~20 hours).

In case of large-scale expression, bacterial cells were harvested by centrifugation (Beckman Coulter Avanti™ J-20 XPI centrifuge) at 4 °C, ≥ 10,000 RCF, for ≥ 15 minutes and the obtained pellet stored at −20 °C until further use. The cell pellet was resuspended with Lysis buffer (Tris-HCl 50 mM, NaCl 500 mM, pH 7.2), supplemented with a cofactor solution of ThDP (final concentration 1 mM) and $MgCl_2$ (final concentration 4 mM) to promote protein folding, also containing 10 mM imidazole if purification was performed. This suspension was subjected to sonication with the MSE Sanyo Soniprep 150 ultrasonic disintegrator, for 10–20 cycles (10 seconds "ON", 10 seconds "OFF") at 15 μA amplitude. The lysate was then centrifuged (Eppendorf Centrifuge 5810 R) at 4 °C, ≥ 18,000 RCF for 30 minutes, to separate the insoluble fraction from the soluble crude lysate (supernatant). The soluble fraction was collected and either used immediately or stored at −80 °C until use.

The purification of holo-split TKs was achieved via Immobilized metal affinity chromatography (IMAC), either by gravity protocol or using a VM20 Vacuum Manifold (Sigma), with Centrifuge Columns (Pierce) loaded with Nickel-charged Chelating Sepharose Fast Flow resin (Cytiva). The purification was carried out using a gradient of imidazole (10 mM, 50 mM, 100 mM) in Tris-HCl 50 mM, NaCl 500 mM pH 7.4, to wash off undesired proteins. Protein elution was

performed with a solution containing imidazole at final concentration of 500 mM and monitored with Bradford Reagent (Bio-Rad). Proteins were concentrated using VivaSpin® 20 centrifugal filters with a 30–50 kDa molecular weight cut-off (Sartorius). Buffer exchange to Tris-HCl 50 mM, NaCl 100 mM pH 7.2 was performed with PD-10 Desalting Columns with Sephadex G-25 resin (Cytiva). Purified proteins were flash frozen in liquid nitrogen and stored at −80 °C.

Protein synthesis and purification fractions were assessed via SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) analysis. 12% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) were used. Samples were resuspended with 4X Laemmli Sample Buffer (Bio-Rad) to a final concentration of 1X and kept at −20 °C until needed. Before loading, samples were incubated 10 minutes at 99 °C. The electrophoretic run was performed in SDS Running buffer 1X, prepared from a 10X stock concentration (for 1 L of 10X stock: 10 g SDS, 144 g glycine, 30.3 g Tris-base), applying a voltage of 200 V for around 40 minutes. PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific) was used as molecular weight reference. Gels were stained with InstantBlue® Coomassie Protein Stain (Abcam).

Protein quantification was initially estimated using Bradford assay. Quick Start Bradford 1X Dye Reagent (Bio-Rad) was used and the protocol for the 250 μL microplate assay followed according to manufacturer. A five point-standard calibration curve with Bovine serum albumin (BSA) was built (0.125–1 mg/mL). Protein concentration was later adjusted based on the limiting subunit through densitometry (ImageJ),[36] in cases where different subunit abundance (N-terminal versus C-terminal) was observed. This adjustment is based on literature evidence for split ChyTK and logical assumption for novel split TKs that the active enzyme is made by both subunits in 1:1 stoichiometry (heterotetramer).

Transketolase Activity Assay

To test if the enzymes produced were active TKs, the TK-synthetic reaction was run with lithium hydroxypyruvate **2** as donor and glycolaldehyde **1** as acceptor for the production of erythrulose **3**. The enzymatic reactions contained 0.05 mg/mL of purified split TK, 2.4 mM ThDP, 9 mM $MgCl_2$, 16 mM **2**, 20 mM **1**, in 50 mM Tris-HCl buffer pH 7.2 (final volume of 150 μL). Reactions were run in 96-well plates for 1 hour at three different temperatures (25 °C, 30 °C, 50 °C), or for 20 hours at 25 °C, shaking at 1000 rpm in a ThermoMixer® C equipped with a ThermoTop (Eppendorf). The change of buffer pH with temperature was accounted for with an estimation factor of −0.028/°C. Enzyme solutions complemented with cofactors were incubated for 20 minutes at 20 °C, 700 rpm before starting the reactions. The reactions were quenched by dilution 1:10 or 1:20 in H_2O containing 0.1% v/v trifluoroacetic acid (TFA), centrifuged 20 minutes at maximum speed (> 20,000 RCF), and the supernatant analyzed *via* HPLC for yield calculation against a standard curve of **3** (Figure S11).

Biochemical Characterization and Determination of Apparent Kinetic Parameters of split TK pQR2834

pQR2834 optima temperature and pH, as well as apparent kinetic parameters were determined by monitoring the initial rates for up to 10 minutes for the production of erythrulose **3** starting from GA **1** and HPA **2**. Experiments were carried out in 1.5 mL-Eppendorf tubes, on a ThermoMixer® C equipped with a ThermoTop (Eppendorf). The enzyme was incubated with cofactors (2.4 mM ThDP, 9 mM $MgCl_2$) for 20 minutes at 20 °C, 700 rpm, prior to the addition of the reaction buffer (50 mM) and the substrates. The change of buffer pH with temperature was accounted for with the appropriate estimation factor. Aliquots were collected at different times and quenched by dilution 1:1 in TFA 0.5% (v/v) in H_2O. Samples were then centrifuged and analyzed by HPLC as previously described. All experiments were performed in duplicate.

For the optimum temperature determination, split TK pQR2834 (0.008 mg/mL) was incubated with 2.4 mM ThDP, 9 mM $MgCl_2$, 10 mM **1**, 10 mM **2**, and Tris-HCl 50 mM pH 7.2, at a range of temperatures (25, 30, 37, 50, and 60 °C), shaking at 750 rpm (final volume of 1 mL). Similar conditions were used for the determination of the optimum pH, but at a fixed temperature of 37 °C and in buffer MES (pH 6.4) or Tris-HCl (pH 7.0, 7.2, 7.8, 8.1, 8.7) (final volume of 0.5 mL).

For the determination of apparent kinetic parameters, the reactions were performed at 37 °C, shaking at 750 rpm, and contained 0.007 mg/mL split TK pQR2834, 2.4 mM ThDP, 9 mM $MgCl_2$, 60 mM of **1**, 1–100 mM of **2**, and Tris-HCl 50 mM pH 7.8 (final volume of 0.4 mL). Data were fitted to the Hill function and apparent kinetic parameters estimated using OriginPro, Version 2022b (OriginLab Corporation, Northampton, MA, USA).

Assessment of Thermal Stability

To evaluate enzyme thermostability, samples of purified split TKs (0.3 mg/mL) were incubated at various temperatures (50–100 °C range) in a ThermoMixer® C equipped with a ThermoTop (Eppendorf) for one hour, followed by rapid cooling on ice. Enzyme activity was then tested (1 hour, 50 °C, 0.05 mg/mL final enzyme concentration) using the previously described TK activity assay method, and HPLC yields were calculated. All experiments were conducted in triplicate.

Native PAGE Analysis

Native PAGE was used to analyze purified split TKs in their native conformation and investigate their oligomeric state. NativeMark™ Unstained Protein Standard (Invitrogen) was used as molecular weight reference. Purified protein samples (1 μg) were incubated for 30 minutes on ice (control), at 25 °C, 50 °C, or 100 °C, in order to investigate the effect of temperature on the formation of oligomers or aggregates. Samples were resuspended with Native PAGE Sample Buffer (Bio-Rad) and run on a 4–20% Mini-PROTEAN® TGX™ Precast Protein Gel. The running buffer was prepared from a concentrated 10X Tris/Glycine Buffer (Bio-Rad).

Analytical HPLC

Analytical high-performance liquid chromatography (HPLC) was performed with Dionex Ultimate™ 3000 HPLC systems (Thermo Fisher Scientific). Reactions with **2** as donor and **1** as acceptor were analyzed using an Aminex HPX-87H column (Bio-Rad) kept at 60 °C as previously described.[4a, g, 6c] An isocratic elution with H_2O with 0.1% v/v TFA at flow rate 0.6 mL/min for 16 minutes was performed. Compounds were detected via UV absorbance (VWD-3400) at 210 nm with following retention times: lithium hydroxypyruvate **2**, RT = 8.6 min; erythrulose **3**, RT = 12.2 min; glycolaldehyde **1**, RT = 12.9 min.

Acknowledgements

The authors would like to acknowledge Prof. Amparo I. Zavaleta (Facultad de Farmacia y Bioquímica, Laboratorio de Biología

Molecular, Universidad Nacional Mayor de San Marcos, Lima, Peru), Dr. Carol N. Flores-Fernández, and Dr. Max Cárdenas-Fernández for their assistance with sampling of Peruvian metagenomes (Maras3, 6Maras and Pilluana). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 956631 (CC-TOP) for A.T. and from the BBSRC Grant BB/X011348/1 for S.A.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: Biocatalysis · Enzymes · Metagenomics · Thermostability · Transketolase

[1] G. Schenk, R. G. Duggleby, P. F. Nixon, *Int. J. Biochem. Cell Biol.* **1998**, *30*, 1297–1318.
[2] a) G. A. Sprenger, U. Schörken, G. Sprenger, H. Sahm, *Eur. J. Biochem.* **1995**, *230*, 525–532; b) J. Littlechild, N. Turner, G. Hobbs, M. Lilly, A. Rawas, H. Watson, *Acta Crystallogr., Sect. D:Biol. Crystallogr.* **1995**, *51*, 1074–1076; c) J. Abdoul-Zabar, I. Sorel, V. Hélaine, F. Charmantray, T. Devamani, D. Yi, V. De Berardinis, D. Louis, P. Marlière, W. D. Fessner, *Adv. Synth. Catal.* **2013**, *355*, 116–128.
[3] H. Yu, Y. Yan, C. Zhang, P. A. Dalby, *Sci. Rep.* **2017**, *7*, 41212.
[4] a) E. G. Hibbert, T. Senussi, S. J. Costelloe, W. Lei, M. E. Smith, J. M. Ward, H. C. Hailes, P. A. Dalby, *J. Biotechnol.* **2007**, *131*, 425–432; b) E. G. Hibbert, T. Senussi, M. E. Smith, S. J. Costelloe, J. M. Ward, H. C. Hailes, P. A. Dalby, *J. Biotechnol.* **2008**, *134*, 240–245; c) P. Payongsri, D. Steadman, J. Strafford, A. MacMurray, H. C. Hailes, P. A. Dalby, *Org. Biomol. Chem.* **2012**, *10*, 9021–9029; d) D. Yi, T. Saravanan, T. Devamani, F. Charmantray, L. Hecquet, W. D. Fessner, *Chem. Commun.* **2015**, *51*, 480–483; e) T. Saravanan, S. Junker, M. Kickstein, S. Hein, M. K. Link, J. Ranglack, S. Witt, M. Lorillière, L. Hecquet, W. D. Fessner, *Angew. Chem. Int. Ed.* **2017**, *56*, 5358–5362; f) T. Saravanan, M.-L. Reif, D. Yi, M. Lorillière, F. Charmantray, L. Hecquet, W.-D. Fessner, *Green Chem.* **2017**, *19*, 481–489; g) H. Yu, R. I. Hernández López, D. Steadman, D. Méndez-Sánchez, S. Higson, A. Cázares-Körner, T. D. Sheppard, J. M. Ward, H. C. Hailes, P. A. Dalby, *FEBS J.* **2020**, *287*, 1758–1776; h) N. Ocal, G. Arbia, A. Lagarde, M. Joly, S. Gittings, K. M. Graham, F. Charmantray, L. Hecquet, *Adv. Synth. Catal.* **2023**, *365*, 78–87; i) A. Mukhopadhyay, K. Karu, P. A. Dalby, *Sci. Rep.* **2024**, *14*, 1287; j) G. Arbia, C. Gadona, H. Casajus, L. Nauton, F. Charmantray, L. Hecquet, *Green Chem.* **2024**, *26*, 7320–7330.
[5] a) M. E. B. Smith, E. G. Hibbert, A. B. Jones, P. A. Dalby, H. C. Hailes, *Adv. Synth. Catal.* **2008**, *350*, 2631–2638; b) C. Zhou, T. Saravanan, M. Lorillière, D. Wei, F. Charmantray, L. Hecquet, W. D. Fessner, D. Yi, *ChemBioChem* **2017**, *18*, 455–459.
[6] a) M. Bawn, F. Subrizi, G. J. Lye, T. D. Sheppard, H. C. Hailes, J. M. Ward, *Enzyme Microb. Technol.* **2018**, *116*, 16–22; b) P. James, M. N. Isupov, S. A. De Rose, C. Sayer, I. S. Cole, J. A. Littlechild, *Front. Microbiol.* **2020**, *11*, 592353; c) M. Cárdenas-Fernández, F. Subrizi, D. Dobrijevic, H. C. Hailes, J. M. Ward, *Org. Biomol. Chem.* **2021**, *19*, 6493–6500; d) A. Yoshihara, Y. Takamatsu, S. Mochizuki, H. Yoshida, R. Masui, K. Izumori, S. Kamitori, *Appl. Microbiol. Biotechnol.* **2023**, *107*, 233–245.
[7] a) J. W. E. Jeffries, N. Dawson, C. Orengo, T. S. Moody, D. J. Quinn, H. C. Hailes, J. M. Ward, *ChemistrySelect* **2016**, *1*, 2217–2220; b) M. S. Carter, X. Zhang, H. Huang, J. T. Bouvier, B. S. Francisco, M. W. Vetting, N. Al-Obaidi, J. B. Bonanno, A. Ghosh, R. G. Zallot, H. M. Andersen, S. C. Almo, J. A. Gerlt, *Nat. Chem. Biol.* **2018**, *14*, 696–705; c) S. Devendran, S. M. Mythen, J. M. Ridlon, *J. Lipid. Res.* **2018**, *59*, 1005–1014; d) L. K. Ly, J. L. Rowles, H. M. Paul, J. M. P. Alves, C. Yemm, P. M. Wolf, S. Devendran, M. E. Hudson, D. J. Morris, J. W. Erdman, J. M. Ridlon, *J. Steroid Biochem. Mol. Biol.* **2020**, *199*, 105567; e) J. A. Shaw, C. A. Henard, L. Liu, L. M. Dieckman, A. Vázquez-Torres, T. J. Bourret, *J. Biol. Chem.* **2018**, *293*, 11271–11282; f) J. Liu, Y. Wei, K. Ma, J. An, X. Liu, Y. Liu, E. L. Ang, H. Zhao, Y. Zhang, *ACS Catal.* **2021**, *11*, 14740–14750; g) R. Chu, Y. Wei, J. Liu, B. Li, J. Zhang, Y. Zhou, Y. Du, Y. Zhang, *Appl. Environ. Microbiol.* **2023**, *89*, e0061723; h) Y. Zhu, T. Shiraishi, J. Lin, K. Inaba, A. Ito, Y. Ogura, M. Nishiyama, T. Kuzuyama, *J. Am. Chem. Soc.* **2022**, *144*, 16715–16719; i) L. Bown, R. Hirota, M. N. Goettge, J. Cui, D. T. Krist, L. Zhu, C. Giurgiu, W. A. van der Donk, K. S. Ju, W. W. Metcalf, *J. Bacteriol.* **2023**, *205*, e0048522.
[8] B. N. Hogg, C. Schnepel, J. D. Finnigan, S. J. Charnock, M. A. Hayes, N. J. Turner, *Angew. Chem. Int. Ed.* **2024**, *63*, e202402316.
[9] S. J. Whiting, PhD thesis, University College London (UK), **2004**.
[10] S. A. Kelly, J. Megaw, J. Caswell, C. J. Scott, C. C. R. Allen, T. S. Moody, B. F. Gilmore, *ChemistrySelect* **2017**, *2*, 9783–9791.
[11] S. C. Cary, I. R. McDonald, J. E. Barrett, D. A. Cowan, *Nat. Rev. Microbiol.* **2010**, *8*, 129–138.
[12] A. Rizzo, C. Aranda, J. Galman, A. Alcasabas, A. Pandya, A. Bornadel, B. Costa, H. C. Hailes, J. M. Ward, J. W. E. Jeffries, B. Dominguez, *ChemBioChem* **2024**, *25*, e202400278.
[13] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, *Nucleic Acids Res.* **2021**, *49*, D412–D419.
[14] a) D. Baud, J. W. E. Jeffries, T. S. Moody, J. M. Ward, H. C. Hailes, *Green Chem.* **2017**, *19*, 1134–1143; b) L. Leipold, D. Dobrijevic, J. W. E. Jeffries, M. Bawn, T. S. Moody, J. M. Ward, H. C. Hailes, *Green Chem.* **2019**, *21*, 75–86.
[15] D. Dobrijevic, L. Benhamou, A. E. Aliev, D. Méndez-Sánchez, N. Dawson, D. Baud, N. Tappertzhofen, T. S. Moody, C. A. Orengo, H. C. Hailes, J. M. Ward, *RSC Adv.* **2019**, *9*, 36608–36614.
[16] S. A. Newgas, J. W. E. Jeffries, T. S. Moody, J. M. Ward, H. C. Hailes, *Adv. Synth. Catal.* **2021**, *363*, 3044–3052.
[17] M. J. S. Bommer, PhD thesis, University College London (UK), **2007**.
[18] R. G. Duggleby, *Acc. Chem. Res.* **2006**, *39*, 550–557.
[19] S. R. Marsden, L. Gjonaj, S. J. Eustace, U. Hanefeld, *ChemCatChem* **2017**, *9*, 1808–1814.
[20] D. Yi, T. Devamani, J. Abdoul-Zabar, F. Charmantray, V. Helaine, L. Hecquet, W. D. Fessner, *ChemBioChem* **2012**, *13*, 2290–2300.
[21] R. Huber, T. A. Langworthy, H. König, M. Thomm, C. R. Woese, U. B. Sleytr, K. O. Stetter, *Arch. Microbiol.* **1986**, *144*, 324–333.
[22] W. J. Jones, J. A. Leigh, F. Mayer, C. R. Woese, R. S. Wolfe, *Arch. Microbiol.* **1983**, *136*, 254–261.
[23] E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, S. T. Sherry, *Nucleic Acids Res.* **2022**, *50*, D20–D26.
[24] a) T. U. Consortium, *Nucleic Acids Res.* **2023**, *51*, D523–D531; b) D. H. Parks, M. Chuvochina, C. Rinke, A. J. Mussig, P. A. Chaumeil, P. Hugenholtz, *Nucleic Acids Res.* **2022**, *50*, D785–D794.
[25] P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, *Bioinformatics* **2014**, *30*, 1236–1240.
[26] F. Madeira, N. Madhusoodanan, J. Lee, A. Eusebi, A. Niewielska, A. R. N. Tivey, R. Lopez, S. Butcher, *Nucleic acids Res.* **2024**, *52*, W521–W525.
[27] R. D. Finn, J. Clements, S. R. Eddy, *Nucleic Acids Res.* **2011**, *39*, W29–37.
[28] M. Steinegger, J. Söding, *Nat. Biotechnol.* **2017**, *35*, 1026–1028.
[29] A. Zielezinski, pfam_scan, Adam Mickiewicz University, Poznań (Poland), **2022**, .
[30] K. Katoh, K. Misawa, K. Kuma, T. Miyata, *Nucleic Acids Res.* **2002**, *30*, 3059–3066.
[31] a) L. Zimmermann, A. Stephens, S. Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *J. Mol. Biol.* **2018**, *430*, 2237–2243; b) F. Gabler, S. Z. Nam, S. Till, M. Mirdita, M. Steinegger, J. Söding, A. N. Lupas, V. Alva, *Curr. Protoc. Bioinf.* **2020**, *72*, e108.
[32] A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, *Bioinformatics* **2009**, *25*, 1189–1191.
[33] S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, *Bioinformatics* **2009**, *25*, 1972–1973.
[34] a) B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, *Mol. Biol. Evol.* **2020**, *37*, 1530–1534; b) S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S.

Jermiin, *Nat. Methods* **2017**, *14*, 587–589; c) D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, *Mol. Biol. Evol.* **2018**, *35*, 518–522.
[35] G. Bianchini, P. Sánchez-Baracaldo, *Ecol. Evol.* **2024**, *14*, e10873.
[36] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, *Nat. Methods* **2012**, *9*, 671–675.
[37] a) X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320–324; b) J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589; c) M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M. Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A. Zidek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis, S. Velankar, *Nucleic Acids Res.* **2024**, *52*, D368–D375; d) I. A. Rodionova, C. Yang, X. Li, O. V. Kurnasov, A. A. Best, A. L. Osterman, D. A. Rodionov, *J. Bacteriol.* **2012**, *194*, 5552–5563; e) J. Van Der Oost, B. Siebers, in *Archaea: Evolution, Physiology, and Molecular Biology* (Eds.: R. A. Garrett, H.-P. Klenk), **2007**, pp. 247–260; f) C. H. Verhees, S. W. Kengen, J. E. Tuininga, G. J. Schut, M. W. Adams, W. M. De Vos, J. Van Der Oost, *Biochem. J.* **2003**, *375*, 231–246; g) S. Mukherjee, D. Stamatis, C. T. Li, G. Ovchinnikova, J. Bertsch, J. C. Sundaramurthi, M. Kandimalla, P. A. Nicolopoulos, A. Favognano, I. A. Chen, N. C. Kyrpides, T. B. K. Reddy, *Nucleic Acids Res.* **2023**, *51*, D957–D963; h) P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. L. de Hoon, *Bioinformatics* **2009**, *25*, 1422–1423.

Manuscript received: December 12, 2024
Revised manuscript received: January 27, 2025
Accepted manuscript online: January 30, 2025
Version of record online: ■■, ■■

RESEARCH ARTICLE

Several putative split transketolases were retrieved via (meta)genome mining, increasing our understanding of this type of prokaryotic TK, whose domains are located on two separate subunits. Ten split TKs were purified and tested for erythrulose production, with the best-performing enzyme kinetically characterized. Evaluation of their thermostability led to the identification of promising candidates for potential use in biocatalysis.

*A. Tonoli, S. Anselmi, J. M. Ward, H. C. Hailes, J. W. E. Jeffries**

1 – 11

Expanding the Enzymatic Toolbox for Carboligation: Increasing the Diversity of the 'Split' Transketolase Sequence Space