

Fine-tuning the activity and stability of an evolved enzyme active-site through noncanonical amino-acids

Henry C. Wilkinson  and Paul A. Dalby 

Department of Biochemical Engineering, University College London, London, UK

Keywords

active-site enzyme engineering; computational molecular docking; enzyme kinetics; enzyme stability; noncanonical amino acids

Correspondence

Paul A. Dalby, Department of Biochemical Engineering, University College London, London, WC1E 6BT, UK
 E-mail: p.dalby@ucl.ac.uk

(Received 15 April 2020, revised 14 August 2020, accepted 2 September 2020)

doi:10.1111/febs.15560

Site-specific saturation mutagenesis within enzyme active sites can radically alter reaction specificity, though often with a trade-off in stability. Extending saturation mutagenesis with a range of noncanonical amino acids (ncAA) potentially increases the ability to improve activity and stability simultaneously. Previously, an *Escherichia coli* transketolase variant (S385Y/D469T/R520Q) was evolved to accept aromatic aldehydes not converted by wild-type. The aromatic residue Y385 was critical to the new acceptor substrate binding, and so was explored here beyond the natural aromatic residues, to probe side chain structure and electronics effects on enzyme function and stability. A series of five variants introduced decreasing aromatic ring electron density at position 385 in the order *para*-aminophenylalanine (*p*AMF), tyrosine (Y), phenylalanine (F), *para*-cyanophenylalanine (*p*CNF) and *para*-nitrophenylalanine (*p*NTF), and simultaneously modified the hydrogen-bonding potential of the aromatic substituent from accepting to donating. The fine-tuning of residue 385 yielded variants with a 43-fold increase in specific activity for 50 mM 3-HBA and 100% increased k_{cat} (*p*CNF), 290% improvement in K_{m} (*p*NTF), 240% improvement in $k_{\text{cat}}/K_{\text{m}}$ (*p*AMF) and decreased substrate inhibition relative to Y. Structural modelling suggested switching of the ring-substituted functional group, from donating to accepting, stabilised a helix-turn (D259-H261) through an intersubunit H-bond with G262, to give a 7.8 °C increase in the thermal transition mid-point, T_m , and improved packing of *p*AMF. This is one of the first examples in which both catalytic activity and stability are simultaneously improved via site-specific ncAA incorporation into an enzyme active site, and further demonstrates the benefits of expanding designer libraries to include ncAAs.

Introduction

As our understanding and control of biocatalytic reactions improves, organic chemists are increasingly able to employ naturally occurring or engineered enzymes to catalyse otherwise difficult synthetic chemical reactions. Transketolase has considerable potential for asymmetric C-C bond formation by catalysing the

transfer of a two-carbon ketol group from a donor substrate to an aldehyde acceptor substrate [1]. The synthesis of α,α' -dihydroxyketone products provides a versatile backbone as a precursor to ketosugars, chiral aminodiols and other high-value molecules such as fragrances and flavours [2–4], while the stereoselectivity

Abbreviations

3-HBA, 3-hydroxybenzaldehyde; ncAA, noncanonical amino acid; *p*AMF, *para*-aminophenylalanine; *p*CNF, *para*-cyanophenylalanine; *p*NTF, *para*-nitrophenylalanine; TK, transketolase; TPP, thiamine pyrophosphate.

of transketolase is highly appealing when attempting C-C bond formation in order to form more complex biologics with multiple stereocentres such as antibiotics. While inherently reversible, the reaction can be rendered irreversible through use of β -hydroxyppyruvate (HPA) as the donor substrate, thus increasing biocatalytic product yields.

Yeast and *E. coli* transketolases function as a homodimer of apparently structurally identical subunits, which is activated upon TPP binding [5]. Each homodimeric unit has two active sites located at the subunit interface [6,7], and each active site can be occupied by a maximum of one TPP molecule and one divalent cation (M^{2+}) such as Ca^{2+} , Mg^{2+} or Mn^{2+} . Recently, the two-species model of transketolase activation was proposed that described the existence of two transketolase subpopulations within a cellular redox regulatory mechanism. Unmodified, inactive TK_{low} is oxidised to form active TK_{high} in response to oxidative stress[8], most likely through sulfenylation/sulfonylation at Cys157 within the TPP-binding site. The proportion of TK_{high} relative to TK_{low} was correlated with the proportion of singly- (+16 Da) and doubly (+32 Da) oxidised TK, relative to unmodified TK_{low} in mass spectra. These monomeric species could combine to form three dimeric species, TK_{high} - TK_{high} , TK_{low} - TK_{low} and the mixed dimer TK_{high} - TK_{low} .

The natural substrates of transketolase are generally phosphorylated sugars such as ribose-5-phosphate and xylulose-5-phosphate, but transketolase has been engineered to accept a plethora of novel substrates by employing various directed evolution strategies. For example, the substrate specificity of transketolase was successfully shifted from phosphorylated sugars, first towards nonphosphorylated, polar substrates [9], then aliphatic nonphosphorylated and heteroaromatic substrates [10–12], and finally to three aromatic benzaldehyde derivatives; 3-formylbenzoic acid (3-FBA), 4-formylbenzoic acid (4-FBA) and 3-hydroxybenzaldehyde (3-HBA) [13,14]. Crystallographic structure analysis of the latter variant, S385Y/D469T/R520Q (3M), coupled with *in silico* molecular docking of the three benzaldehyde derivatives, revealed the creation of two distinct binding pockets that are sterically separated by the D469T mutation [15]. The S385Y mutation was predicted to play a crucial role in aromatic substrate binding to both pockets through π - π stacking interactions with F434 and the aromatic ring of the substrate. While 3-FBA bound into one pocket and 4-FBA the other, both with relatively high affinity and catalytic turnover, 3-HBA was found to bind to both pockets 1 and 2 with low-affinity and poor catalytic productivity.

In previous rounds of directed evolution, activity gains were realised by semirational engineering of important active-site residues through relatively large saturation mutagenesis libraries, and yet also required re-engineering of protein stability due to critical stability losses incurred by active-site mutations [16]. Indeed, the 3M variant was similarly found to be less stable than the wild-type and has recently been restabilised through additional nonactive-site mutations [17]. The aim of this study was to further explore the critical aromatic ring of Y385, through altered aromatic ring electron density, to probe and potentially improve the catalytic activity, substrate inhibition, enzyme stability and binding pocket preference of the 3M variant simultaneously, without the need for additional restabilising mutations. We then also rationalised the observed trends in the context of active-site electronics, hydrophobic packing, and the size, polarity and hydrogen binding potential of the ring-substituted functional group.

The genetic code limits us to only 20 amino acids. Only phenylalanine has a less electron-dense single aromatic ring than tyrosine, and none are more electron-dense. We therefore expanded the genetic code beyond its natural limits via incorporation of noncanonical aromatic amino acids (ncAAs) to create a series of five variants with highly electron-donating to highly electron-withdrawing aromatic ring substitutions at the *para*-position, (*p*-aminophenylalanine (*p*AMF)> tyrosine (Y)> phenylalanine (F)> *p*-cyanophenylalanine (*p*CNF) and *p*-nitrophenylalanine (*p*NTF), in order of decreasing aromatic ring electron density).

The incorporation efficiency (protein yield relative to wild-type) for various incorporation systems and their respective ncAAs has been well documented and quantified in the majority of studies to date. However, their incorporation fidelity (proportion of incorporated ncAA relative to misincorporated amino acids) is comparatively under-reported. Often, the misincorporation mass spectrometry peaks are hidden by the dominant ncAA-incorporated peak [18,19], or mass spectra are provided but not deconvoluted and/or quantified [18,20]; otherwise, the deconvoluted spectra are poorly resolved due to the low sensitivity of the LCMS [21,22]. In many others, no supporting mass spectra data are provided at all, giving no information on the level of misincorporation in the presence of ncAA [23–28]. Incorporation fidelity is a factor that should not be ignored for enzyme activity, ligand binding or protein stability studies, since heterogeneity is almost always unavoidable in ncAA-incorporation studies, and may hide the true performance of the ncAA-incorporated species. We therefore describe a methodology

to (a) accurately quantify incorporation fidelity from deconvoluted mass spectra and (b) account for the contribution of misincorporated species to the experimentally determined or ‘apparent’ activity and stability parameters, to reveal the ‘true’ parameters of the ncAA-incorporated species of interest. The suitability of the ‘apparent’ and ‘true’ parameters as indicators of biocatalytic output, structure–function relationships and enzymatic potential is also discussed.

Global incorporation of ncAAs has enhanced the activity and thermal stability of a number of enzymes, mainly through incorporation of fluorinated natural amino acid analogues [29–34]. Site-specific, active-site ncAA incorporation is a nascent and promising field of research that was initially held back by low incorporation efficiencies and fidelities of incorporation. Over the last decade, the genetic code has been expanded to incorporate a myriad of ncAAs with much improved efficiencies and fidelities. To date, only a handful of studies have successfully improved [26,27,35] or introduced novel catalytic function [18,28], or enhanced thermostability [19,25] via site-specific ncAA incorporation, and even less introduced them into an active-site. While the latter study [25] observed both a 1% increase in T_m and a modest 15% improvement in catalytic activity at 40 °C relative to wild-type, the thermostable mutant was less-active than the wild-type at 23 °C, which suggested the improved activity reflected only the improved thermostability at 40 °C rather than a genuine improvement in catalysis. To our knowledge, we therefore report the first example of a genetically encoded, site-specific active-site ncAA-incorporated variant with both enhanced activity at 22 °C and thermostability (T_m), and demonstrate the benefits of including ncAAs in site-specific, smart designer libraries.

Results

The electronic properties of para-substituted phenylalanine derivatives

The strength of π – π stacking interactions is influenced by aromatic ring electron density. Electron-withdrawing

groups (e.g. $-\text{NO}_2^-$) strengthen hydrophobic interactions between aromatic rings [36], while the opposite is true for electron-donating groups (e.g. $-\text{NH}_2$). The amino acids used in this study can be ranked in order of aromatic ring electron density from most dense to least dense: *p*-aminophenylalanine (*p*AMF) > tyrosine (Y) > phenylalanine (F) > *p*-cyanophenylalanine (*p*CNF) > *p*-nitrophenylalanine (*p*NTF) (Fig. 1). The aromatic ring electron density at position 385 may therefore influence: (a) active-site hydrophobicity; (b) the overall hydrophobic packing of active-site residues; and (c) the strength of substrate binding due to π – π stacking interactions, in either catalytically productive or inhibitory orientations.

Nevertheless, substrate binding and catalytic turnover are dictated by a multiplicity of additional factors, including hydrogen bonding (H-bonds) with other active-site residues, steric hindrance and orientation of the acceptor substrate relative to the dihydroxyethyl-TPP (DHE-TPP) intermediate in three-dimensional space. Furthermore, the characteristics of the ring-substituted functional groups may also have an impact on both catalysis and stability, as they differ in size, polarity and H-bonding potential from donating (*p*AMF and Y), to nonbonding (F), and finally accepting (*p*CNF and *p*NTF). While it is relatively hard to predict the outcome of such changes without far more structural information and computational power, one can predict with some confidence that one or more properties of an enzyme will at least change when altering the electronics of an active-site residue that is already known to be critical to the acceptance of aromatic substrates.

Efficiency and fidelity of ncAA incorporation

An evolved, orthogonal *M. jannaschii* aminoacyl-tRNA synthetase/tRNA pair [37], which incorporates ring-substituted phenylalanine derivatives at amber stop codons with high incorporation fidelities (i.e. low levels of natural amino acid misincorporation) and efficiencies (i.e. high protein yield) [20], but considerable ncAA promiscuity [22], was coexpressed using an optimised plasmid, *pUltra*, with the TK expression plasmid, *pQR791-TK-S385X/D469T/R520Q*, in *C321.ΔA.exp*. This ‘amberless’

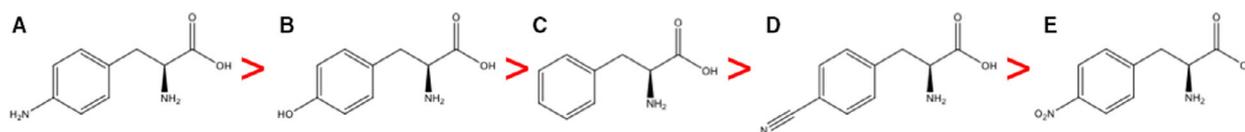


Fig. 1. Chemical structures of the five single-ringed aromatic amino acids utilised in this study. In order of decreasing aromatic ring electron density: *p*-aminophenylalanine (*p*AMF) (A), tyrosine (Y) (B), phenylalanine (F) (C), *p*-cyanophenylalanine (*p*CNF) (D) and *p*-nitrophenylalanine (*p*NTF) (E).

E. coli strain, which has all amber codons replaced, and RF1 deleted, and thus cannot terminate translation at amber stop codons, giving an improved absolute efficiency of ncAA incorporation by preventing the formation of truncated protein [38]. The TK 3M variants generated are henceforth designated as 385*p*AMF, 385Y, 385F, 385*p*CNF and 385*p*NTF.

A methodology to quantify incorporation fidelity from deconvoluted mass spectra

The major drawback of using expression strains such as *BL21 (DE3)* for ncAA incorporation is the potential to produce truncated protein rather than full-length protein. Protein expression in *C321.ΔA.exp* prevented translation of truncated protein and significantly improved the efficiency of multisite ncAA incorporation [39]. However, only marginal improvements were observed in single-site incorporation [38] and it is still unclear what impact strain usage may have on the fidelity of incorporation. We therefore compared the incorporation into N-terminal His-tagged TK using the pUltra plasmid system in both *C321.ΔA.exp* and *BL21(DE3)*, to choose the best expression strain. We developed a methodology to quantify the incorporation fidelity based on the hierarchy of TK species and subspecies outlined below (Fig. 2).

- 1 (nc)AA – the target ncAA (*p*AMF, *p*CNF and *p*NTF) or natural amino acid (AA) (Y and F).
- 2 TK_{biocat} – all TK species used in biocatalysis or present in biocatalytic samples. Includes TK_{incorporated}, TK_{misincorporated} and their subspecies.
- 3 TK_{incorporated} – all TK species with the target (nc) AA incorporated at residue 385. Includes all TK_{low},

TK_{active} and TK_{inactive} subspecies of the target (nc) AA.

- 4 TK_{misincorporated} – all TK species with Gln, Phe or Tyr misincorporated at residue 385. Includes all TK_{low} and TK_{active} subspecies. The TK_{inactive} subspecies of TK_{misincorporated} was negligible.
- 5 TK_{low} – unmodified (+0 Da), low-activity TK subspecies.
- 6 TK_{active} – singly (+16 Da) or doubly (+32 Da) oxidised, active TK subspecies (i.e. TK_{high}).
- 7 TK_{inactive} – overoxidised (+48 Da, +64 Da, +80 Da etc.), inactive TK subspecies.
- 8 % (nc)AA_{incorporated} – proportion of (nc)AA-TK_{incorporated} relative to all TK species.
- 9 % (nc)AA_{active} – proportion of all TK_{high} subspecies of (nc)AA_{incorporated} relative to all TK species.

The incorporation fidelity of each ncAA was quantified by fitting triplicate mass spectra to the sum of multiple Gaussian functions in OriginLabs (Fig. 3). After assigning initial values by eye, the peak width parameter was shared between the summed Gaussian functions, while the peak area, peak width and peak centre parameters were unconstrained in order to converge towards their best fit values.

TK was recently shown to exist in multiple oxidation states [8], and therefore, each TK_{misincorporated} and TK_{incorporated} species consisted of a series of subspecies of unmodified, low-activity TK_{low} (peak 1 = +0 Da); two oxidised, active TK_{high} species (collectively called TK_{active} from henceforth) (peak 2 = +16 Da; peak 3 = +32 Da) and overoxidised, inactive TK_{inactive} (peak 3 = +48 Da, peak 4 = +64 Da, peak 5 = +80 Da, etc.) (Fig. 3). The two TK_{high} peaks of each misincorporated species were small and undetectable by eye in

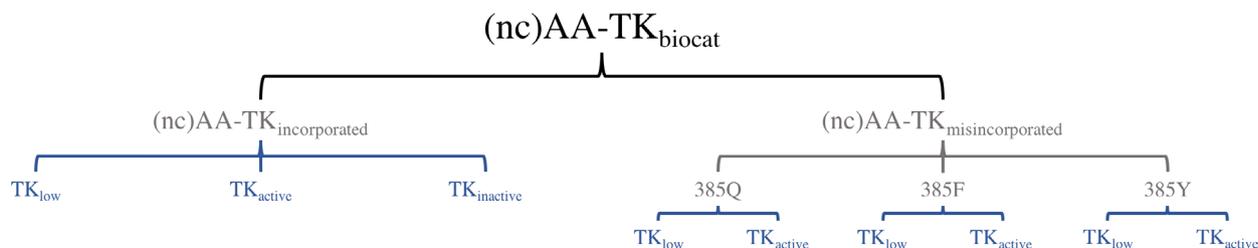


Fig. 2. A schematic diagram outlining the hierarchy of TK species (grey) and subspecies (blue). To provide clarity, the following example is given for variant *p*AMF. *p*AMF-TK_{biocat} (black) is defined as all *p*AMF species and subspecies contained in any biocatalytic sample. The sample can be subdivided into four species (grey): one *p*AMF-incorporated (*p*AMF-TK_{incorporated}) species and three *p*AMF-misincorporated (*p*AMF-TK_{misincorporated}) species (385Q, 385F and 385Y), where the natural amino acids glutamine (Q), phenylalanine (F) or tyrosine (Y) have been misincorporated at residue 385. Each of these four species can be subdivided into three subspecies (blue): low-activity TK (TK_{low}), active TK (TK_{active}) and inactive TK (TK_{inactive}). However, the proportion of variant *p*AMF that existed as the TK_{inactive} subspecies of 385Q, 385F and 385Y was negligible and hence excluded from all calculations and from Figure 2.

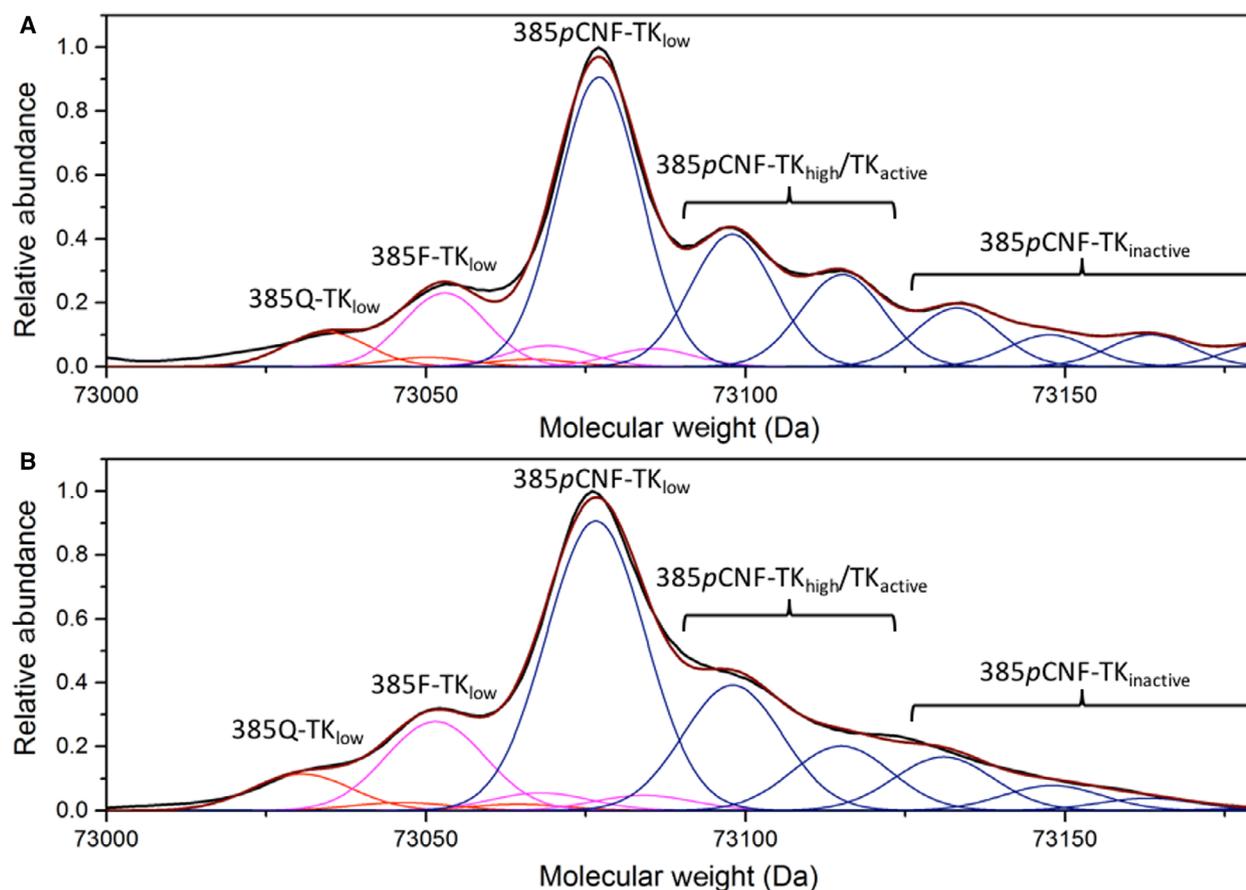


Fig. 3. Mass spectra of variant 385pCNF when expressed in (A) *C321.ΔA.exp* or (B) BL21 (DE3). Each mass spectrum ($n = 3$; average spectra in black) was fitted to the sum of multiple Gaussian functions (smaller individual peaks) corresponding to unmodified (TK_{low}), single- and double-oxidised forms (TK_{high} or $\text{TK}_{\text{active}}$) of the misincorporated species 385Q (red) and 385F (magenta); and TK_{low} , TK_{high} and the overoxidised, inactive forms ($\text{TK}_{\text{inactive}}$) of the ncAA-incorporated species 385pCNF (blue). These same species have been observed previously in both wild-type TK [8] and variant S385Y/D469T/R520Q [40]. The ratio of the peak area of unmodified, singly- and doubly oxidised peaks of Y and F was determined from their mass spectra (Fig. S1B,C), and applied to the mass spectra of 385pCNF. The cumulative fit of all peaks is shown in wine-red. The proportion of active 385pCNF relative to all TK species (both incorporated and misincorporated), $\%p\text{CNF}_{\text{active}}$, was determined from the peak area of singly- and doubly oxidised 385pCNF- TK_{high} relative to the cumulative peak area (Table 1). The $\% \text{TK}_{\text{incorporated}}$, $\% \text{TK}_{\text{misincorporated}}$ and $\% \text{TK}_{\text{active}}$ were determined for 385pAMF, 385Y, 385F and 385pNTF (Table S1) from their respective mass spectra (Fig. S1A,D) using the same methodology. The expected and observed molecular weights are listed in the Supporting Information (Table S2).

Table 1. The $\% \text{TK}_{\text{incorporated}}$ ($\% p\text{CNF}_{\text{incorporated}}$), $\% \text{TK}_{\text{misincorporated}}$ ($\% F_{\text{misincorporated}}$ and $\% Q_{\text{misincorporated}}$) and $\% p\text{CNF}_{\text{active}}$ when expressed in (A) *C321.ΔA.exp* or (B) BL21 (DE3). Parameters and their associated SEM are from triplicate mass spectra ($n = 3$) (Fig. 3).

Strain	$\% p\text{CNF}_{\text{incorporated}}$	$\% F_{\text{misincorporated}}$	$\% Q_{\text{misincorporated}}$	$\% p\text{CNF}_{\text{active}}$
<i>C321.ΔA.exp</i>	$76.8 \pm 4.0\%$	$16.0 \pm 2.4\%$	$7.4 \pm 0.9\%$	$22.0 \pm 4.1\%$
BL21 (DE3)	$73.8 \pm 1.9\%$	$18.5 \pm 1.2\%$	$7.6 \pm 0.7\%$	$24.3 \pm 1.0\%$

each mass spectra since they were dwarfed by the larger TK_{low} peaks. However, formation of TK_{high} is a redox-regulated cellular process and the ratio of $\text{TK}_{\text{high}}:\text{TK}_{\text{low}}$ is relatively constant but variant-specific [40]. We therefore assumed that a) the two TK_{high}

peaks of each $\text{TK}_{\text{misincorporated}}$ species were present and still contributed towards the overall peak area of the mass spectra; and b) the ratio of $\text{TK}_{\text{high}}:\text{TK}_{\text{low}}$ was equivalent in naturally incorporated and misincorporated natural amino acid species.

We estimated the peak area of the two TK_{high} peaks of each $\text{TK}_{\text{misincorporated}}$ species by applying the peak area ratio of peak1:peak2 and peak1:peak3, obtained from the mass spectra (Fig. S1B-C), to peak1 of each $\text{TK}_{\text{misincorporated}}$ species within the mass spectra of each ncAA- $\text{TK}_{\text{incorporated}}$ variant (Fig 3; Fig. S1A & D). Therefore, the peak centre and peak area of the two oxidised TK_{high} peaks of each misincorporated subspecies were fixed and the function refitted to obtain peak areas for peaks 1–3 of each misincorporated subspecies and peaks 1–7 of the ncAA-incorporated species within a single mass spectra. The fidelity of incorporation and the terms defined above were subsequently determined from the peak areas derived from the mass spectra of

each ncAA variant, and were also applied to the experimentally determined activity and stability data below.

Strain choice has negligible impact on incorporation fidelity

Comparison of the mass spectra of 385pCNF when coexpressed with pUltra in either *C321.ΔA.exp* or *BL21 (DE3)* confirmed that in the case of TK, strain choice made little difference to the incorporation fidelity ($76.8 \pm 4.0\%$ and $73.8 \pm 1.9\%$, respectively) (Fig. 3; Table 1). Since there was negligible difference between the incorporation fidelity of the two expression strains, all subsequent variants were expressed in

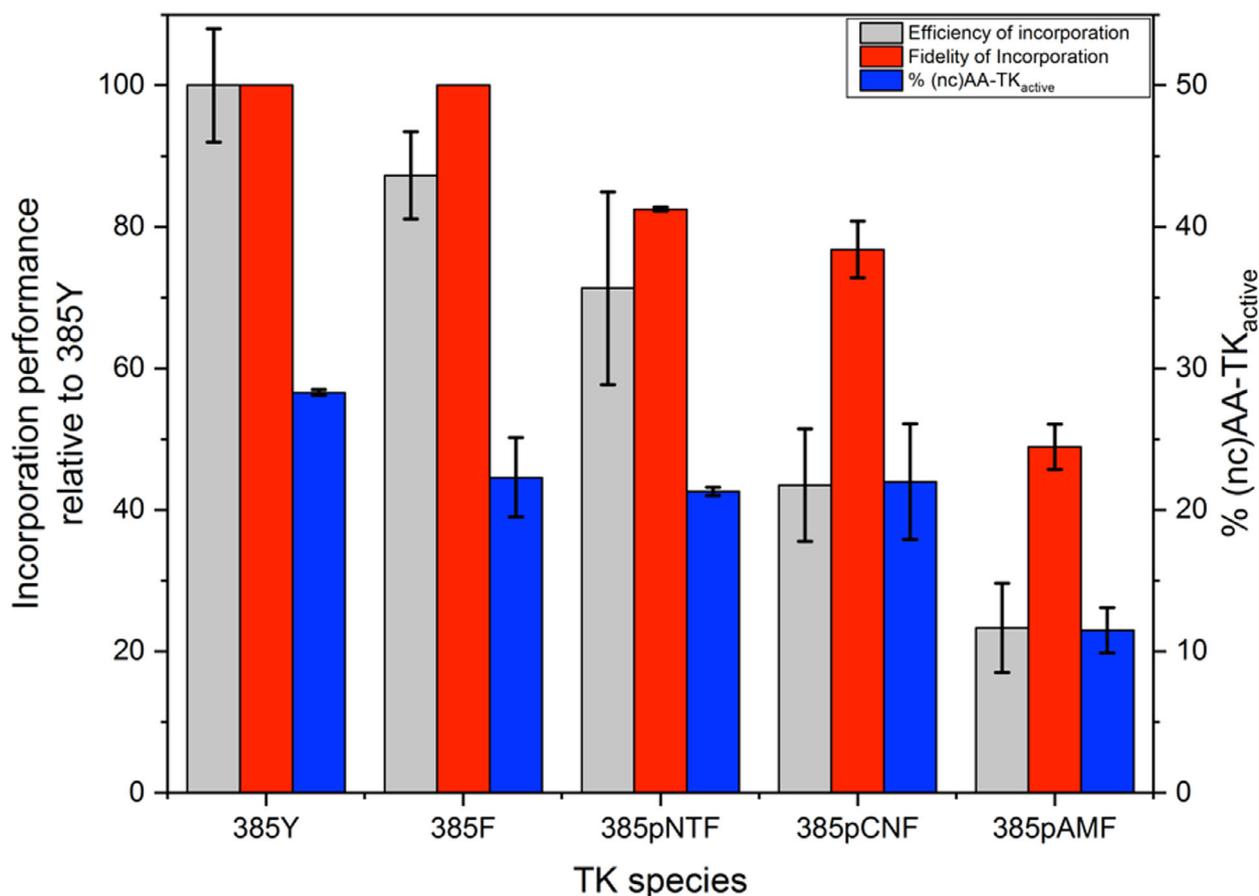


Fig. 4. Incorporation efficiency (grey), fidelity (red) relative to 385Y and the $\% \text{TK}_{\text{active}}$ (blue) at residue 385. The incorporation efficiency, defined as the relative yield of TK (incorporated and misincorporated) relative to 385Y, was determined from densitometry analysis of SDS/PAGE ($n = 3$) (Fig. 2). The incorporation fidelity, defined as the proportion of ncAA-incorporated species relative to all misincorporated species (385Q, 385F and 385Y), was determined from analysis of mass spectra ($n = 3$) (Fig. 3; Table 1; Fig. 1). The proportion of active $\text{TK}_{\text{incorporated}}$ relative to all TK species (both incorporated and misincorporated), $\% \text{TK}_{\text{active}}$, was determined from the peak areas of the singly- and doubly oxidised TK_{high} species relative to the cumulative peak area (Fig. 3; Table 1; Fig 1). Error bars represent the standard error of the mean (SEM).

C321.ΔA.exp to avoid purification of any truncated protein when using an N-terminal His-tag.

The fidelity and efficiency of incorporation is ncAA-dependent

Figure 4 summarises the incorporation efficiency and fidelity as observed by SDS/PAGE and LC-ESI-MS, respectively (for full analysis and the supporting SDS/PAGE gels and deconvoluted mass spectra, see SI Text 1, SI Fig. S1 and SI Fig. S2). In all ncAA variants, we observed a low but significant level of misincorporation of glutamine, phenylalanine and sometimes tyrosine, in addition to the TK_{low}, TK_{high} and TK_{inactive} forms [8] of both ncAA-incorporated and misincorporated species. The proportions determined by mass spectrometry were subsequently used to extract the catalytic and stability parameters of the ncAA-incorporated species from our enzyme kinetic measurements. Overall, the *M. jannaschii* aminoacyl-tRNA synthetase/tRNA incorporation system was found to prefer ring-substituted phenylalanine derivatives with strong electron-withdrawing groups at the para position over those with electron-donating groups.

How heterogeneous is too heterogeneous?

The existence of multiple TK species in any given sample raised intriguing questions that were relevant not only to this study but a large number of proteins. Nature has evolved complex regulatory mechanisms and networks to facilitate dynamic intracellular responses to exploit or survive changes in external conditions. Many regulated proteins will therefore exist in more than one noncovalently or covalently modified state, which can be in reversible and/or irreversible equilibria. Other protein modifications are an often-unavoidable result of cellular conditions, such as the overoxidation of solvent-exposed cysteine residues. In most studies, these species cannot, and are not, readily separated during purification. The same is true in ncAA incorporation, where heterogeneous samples of ncAA-incorporated and misincorporated natural amino acid species are frequently, if not always, observed. Where we should draw the line between homogeneous vs heterogeneous samples is nontrivial, but ultimately depends on the context. Industrial uses of enzymes in biocatalysis and proteins as therapeutics evaluate the overall performance of the heterogeneous mixture in processes or clinical trials. However, understanding and potentially controlling the underlying heterogeneity and its impact on performance is still important. For example, we now know that TK exists

as multiple homodimeric species [8,40], and yet all previous work evaluating the performance of TK variants in biocatalytic processes is no less useful or valid. Instead, the performance measured characterises the ‘apparent’ activity and stability of that variant as a biocatalyst, including all unmodified and modified species within the sample, obtained using a particular production process.

However, to increase our understanding of the relative contributions of subspecies, and to potentially modify these for improving biocatalysis, there is value in attempting to deconvolve the ‘true’ kinetic parameters for individual species, from the total ‘apparent’ kinetic parameters, using simultaneous equations. We therefore aimed to deconvolve the relative contributions from natural amino acid misincorporation during ncAA incorporation, to reveal variants with significantly improved ‘true’ parameters, that would otherwise be overlooked.

The catalytic performance of 385pAMF, 385Y, 385F, 385pCNF and 385pNTF

Previously, the K_m and k_{cat} of the 3M triple-mutant towards 3-HBA were reported as 390 ± 10 mM and 2.1 ± 0.2 s⁻¹, respectively, using a substrate concentration up to 25 mM 3-HBA owing to insolubility at higher concentrations [14]. As a result, the reported K_m and k_{cat} were overestimated because saturating concentrations of substrate were never reached. By sonication and careful pH adjustment, we have now improved the solubility limit of 3-HBA to at least 50 mM, allowing a more accurate calculation of the true kinetic parameters. Initially, the specific activity of each variant was determined at 50 mM 3-HBA. The catalytic performance of each variant was then investigated further over a range of [3-HBA] (Fig. 5).

The sliding-substrate inhibition Model

The significant substrate inhibition observed at above 25 mM for all variants indicated the binding of multiple substrate molecules or inhibitory poses, given the ‘cooperative’ shape of the curve at high [S] (Fig. 5). The experimental data were first fitted to the standard Michaelis–Menten-derived function describing a sequential-order noncompetitive inhibition (NCI) model (Fig. 5A; Equation 1, Materials and Methods). In this model, enzyme activity decreased when the enzyme was bound to a single inhibitor molecule (in this case also the substrate) regardless of whether the enzyme was bound to the substrate or inhibitor first. With the exception of 385pNTF (Fig. 5A, magenta),

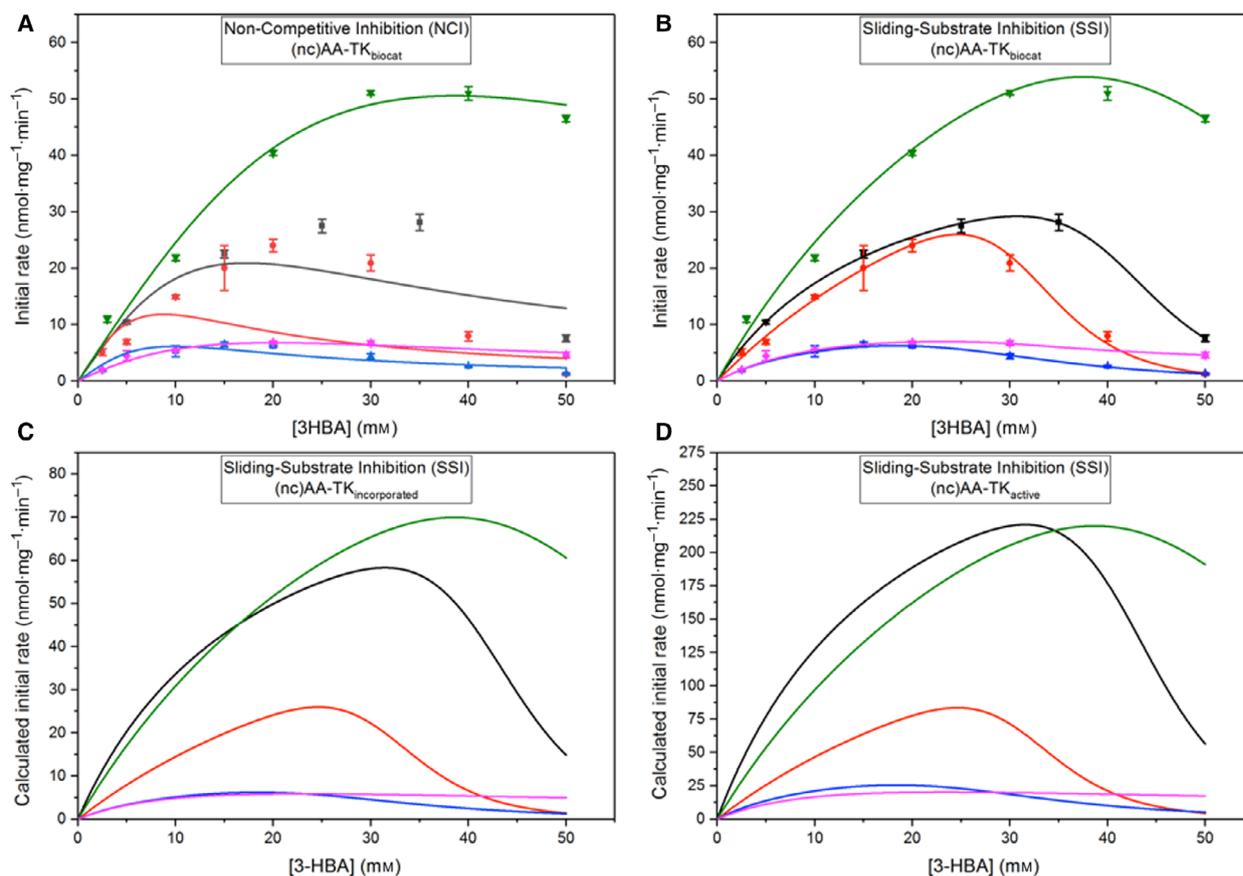


Fig. 5. The apparent and true initial rates of variants 385pAMF (black), 385Y (red), 385F (blue), 385pCNF (green) and 385pNTF (magenta) towards 3-HBA. The initial rate of each variant was determined at between 0–50 mM 3-HBA, 9 mM Mg^{2+} and 2.4 mM TPP and fitted to (A) the noncompetitive inhibition (NCI) model and (B–D) the sliding-substrate inhibition (SSI) model (derived in Materials and Methods). In (A) and (B), the same experimental data were fitted to the NCI and SSI model, respectively, to demonstrate the superior fit to the SSI model. (C) and (D) visualise the predicted initial rate curves by populating the SSI model with the kinetic parameters of $TK_{incorporated}$ and TK_{active} species, respectively, which were in turn determined from the experimental data. Measurements were taken in triplicate ($n = 3$), and the error bars represent the standard error of the mean (SEM).

the function gave an inadequate fit to the experimental data, which implied the observed acceptor substrate inhibition could not always be explained by the standard noncompetitive inhibition model alone.

Subsequently, a modified Michaelis–Menten function, the sliding-substrate inhibition (SSI) model, was derived (Equation 2, *Materials and Methods*) that describes an initial catalytically productive enzyme-substrate binding event with one substrate molecule, $n_i = 1$, that is completely inhibited at high $[S]$ in a second collective binding event, but allowing for multiple inhibitory enzyme-substrate interactions, $n_i \geq 1$. The function assumed that binding is sequential, with the catalytically productive substrate binding preceding subsequent inhibitory binding. The inhibitory poses are likely to have different binding constants (K_i) and so in theory, the substrate could ‘slide’ between different inhibitory binding poses.

‘Sliding’ is significantly more likely to be observed in catalytic reactions that are dominated by hydrophobic enzyme-substrate interactions, such as the interactions between the hydrophobic S385Y/D469T/R520Q active site and 3-HBA, which result in much less-well defined enzyme-substrate orientations compared to electrostatic and hydrogen-bonded interactions. For simplicity, the n_i number of inhibitory poses were assumed to be represented by a single average inhibitory dissociation constant, K_i .

The experimental data of all variants fitted extremely well to the SSI model (Fig. 5B) and were therefore considered to be a good kinetic model for TK. While the kinetic data of 385pNTF also fitted well to the SSI model, we observed a superior fit to the NCI model, which was equivalent to fixing $n = 2$ (and $n_i = 1$) in the SSI model (Equation 1, *Materials and Methods*).

The n_i value was fixed at 1 in the SSI model when determining the kinetic parameters of 385pNTF, while the SSI model was used for all other variants. The experimental data of each variant fitted equally well with identical parameters to the randomly ordered SSI model (Fig. S3, Equation S1), below, which indicated that the additional random-order term did not contribute to the overall fit, thus validating the prior assumption that the reaction was sequential in order.

Determination of kinetic parameters for (nc)AA-TK_{biocat}, (nc)AA-TK_{incorporated} and (nc)AA-TK_{active}

The analysed mass spectrometry data were combined with the experimentally determined activity data and then globally fitted to the weighted sum of three SSI model functions (Equation 3, Materials and Methods).

For each variant, the apparent kinetic parameters derived from the experimental data were a convolution of the contributions from the various (nc)AA-TK_{incorporated} and (nc)AA-TK_{misincorporated} subspecies (Fig. 2). By contrast, the kinetic data for natural incorporation of F and Y at residue 385 were not convoluted and so were obtained directly, thus also allowing their contributions to the other variant kinetics to be accounted for. Three variations of this function were used to analyse the kinetic parameters of (nc)AA-TK_{biocat}, (nc)AA-TK_{incorporated} and (nc)AA-TK_{active} of each TK variant (Table 2). The pros and cons of each method are discussed in turn below.

Analysis 1: No deconvolution (TK_{biocat})

The first analysis compared the kinetic parameters of each transketolase variant as an overall biocatalytic sample, TK_{biocat}, without attempting to deconvolve the parameters of each species and subspecies. The activity data were fitted to a single SSI function (Fig. 5B).

Analysis 1 provided information on the performance of each variant (and all subpopulations) as an overall biocatalyst. However, it was the highest level analysis and provided only 'apparent' activity and stability parameters of TK_{biocat}, but did not reflect the 'true' catalytic performance of the ncAA-incorporated subpopulations, especially in cases where the ncAA is incorporated with low fidelity (i.e. high % misincorporation). This analysis indicated 385pCNF to be the best overall biocatalyst. Incorporation of pCNF into the active site of S385pCNF/D469T/R520Q improved the k_{cat} , K_i , n_i and k_{cat}/K_m by 57%, 74%, 57% and 64%, respectively, relative to variant 385Y. The K_m of 385pCNF was within error of that for variant 385Y. Incorporation of pAMF at residue 385 gave a 140% improvement in K_m but decreased the k_{cat} by 38%, resulting in a 45% improvement in catalytic efficiency (k_{cat}/K_m) relative to 385Y. Finally, incorporation of F and pNTF at residue 385 significantly improved their K_m , while the catalytic turnover was severely impaired, resulting in overall lower catalytic efficiencies than for 385Y.

Table 2. The kinetic parameters of the TK_{biocat}, TK_{incorporated} and TK_{active} species and subspecies of variants 385pAMF, 385Y, 385F, 385pCNF and 385pAMF towards 3-HBA. Associated errors are the fitting error for the weighted sum of three sliding-substrate inhibition (SSI) model functions. The %(nc)AA_{incorporated}, %(nc)AA_{misincorporated} and %(nc)AA_{active} used to weight the function were determined from the peak areas from LC-ESI-MS data as described in Fig. 3 and the Materials and Methods. Note that 385pNTF-TK_{incorporated} and 385pNTF-TK_{active} fitted best to the NCI model; hence, $n_i = 1$ and has no associated error.

Variant	Rel. spec activity (50 mM)	K_m (mM)	k_{cat} (min ⁻¹)	K_i (mM)	n_i	k_{cat}/K_m (s ⁻¹ M ⁻¹)
385pAMF-TK _{biocat}	5.5 ± 0.6	18.7 ± 1.7	3.6 ± 0.2	41.3 ± 1.0	8.6 ± 0.9	3.2 ± 0.2
385pAMF-TK _{incorporated}	6.5 ± 0.7		7.1 ± 0.4	42.1 ± 1.9	9.5 ± 4.0	6.3 ± 0.5
385pAMF-TK _{active}	12.8 ± 1.5		26.7 ± 1.0	42.2 ± 2.1	9.6 ± 2.8	23.8 ± 0.8
385Y-TK _{biocat}	–	44.8 ± 16.3	5.8 ± 0.2	29.7 ± 1.1	7.7 ± 0.7	2.2 ± 0.6
385Y-TK _{incorporated}	–					
385Y-TK _{active}	–		18.7 ± 0.5			6.9 ± 1.1
385F-TK _{biocat}	1.0 ± 0.1	11.6 ± 2.2	0.8 ± 0.1	29.4 ± 1.9	3.8 ± 0.3	1.2 ± 0.2
385F-TK _{incorporated}	0.9 ± 0.1					
385F-TK _{active}	1.2 ± 0.1		3.4 ± 0.1			4.8 ± 0.1
385pCNF-TK _{biocat}	33.5 ± 3.3	41.8 ± 12.8	9.1 ± 0.3	51.7 ± 1.8	4.9 ± 2.5	3.6 ± 0.8
385pCNF-TK _{incorporated}	26.4 ± 2.6		11.7 ± 0.3	52.1 ± 1.3	5.3 ± 2.1	4.7 ± 1.0
385pCNF-TK _{active}	42.7 ± 4.2		36.7 ± 0.6	52.2 ± 0.9	5.3 ± 1.4	14.7 ± 0.2
385pNTF-TK _{biocat}	3.4 ± 0.4	14.8 ± 4.2	1.0 ± 0.2	38.1 ± 8.5	2.0 ± 0.6	1.2 ± 0.3
385pNTF-TK _{incorporated}	2.3 ± 0.3		1.0 ± 0.1	35.8 ± 20.1	1.0 ± –	1.1 ± 0.2
385pNTF-TK _{active}	3.9 ± 0.5		3.4 ± 0.3	37.1 ± 14.1		3.8 ± 0.6

Analysis 2: Deconvolution of ncAA-TK_{incorporated} species parameters

The ‘true’ kinetic parameters of each (nc)AA-TK_{incorporated} variant were deconvoluted from the experimentally determined ‘apparent’ data, by accounting for the presence of (nc)AA-TK_{misincorporated} species, and their contribution towards the total activity/stability (Fig. 5C). The three SSI functions were weighted to the %ncAA_{incorporation} %F_{misincorporation} and %Y_{misincorporation} determined from the mass spectra of each variant (Fig. 3; Equation 3, Materials and Methods; Fig. 1, supplementary information). The ‘true’ catalytic parameters of naturally incorporated 385F-TK_{incorporated} and 385Y-TK_{incorporated} were determined from experimental activity data (Fig. 5; Table 2). The specific activity of 385Q at 50 mM 3-HBA was negligible relative to 385Y, and most variants had a small %Q_{misincorporation}, and therefore, the contribution of 385Q towards overall activity was negligible. Two of the three weighted SSI model functions were populated with the catalytic parameters of 385F-TK_{incorporated} and 385Y-TK_{incorporated}. The K_m of the third SSI function, corresponding to the ncAA-TK_{incorporated} species, was fixed at the TK_{biocat} K_m value, while the other parameters were floated to extract the ‘true’ kinetic parameters of the ncAA-TK_{incorporated} species of 385pAMF, 385pCNF and 385pNTF (visualised in Fig. 5C).

Analysis 2 had the largest relative impact on 385pAMF because a high level of misincorporation was observed in this variant. Consequently, after accounting for misincorporation, the incorporation of pAMF at residue 385 improved the k_{cat} by 22% relative to 385Y, rather than the 38% decrease observed in Analysis 1 (Table 2). This improvement in k_{cat} also resulted in an impressive 186% improvement in the k_{cat}/K_m relative to 385Y. Thus, strategies to minimise misincorporation in 385pAMF would be greatly beneficial for this variant as a biocatalyst. The major caveat to Analysis 2 is that it averages out effects that may have arisen from any population of mixed dimers between the monomeric subunits of different subspecies. It also ignores any potential influence of the variant upon the ratio of TK_{high}:TK_{low} [40], which could in turn affect the overall activity, as well as the inhibition of TK_{high} by TK_{low} in the resulting mixed dimer species.

Analysis 3: Deconvolution of (nc)AA-TK_{active} subspecies parameters

Analysis 3 provided the most detailed information about the ‘true’ parameters of only the active

subspecies within each (nc)AA-incorporated TK variant, (nc)AA-TK_{active} (visualised in Fig. 5C). The same methodology was employed as in Analysis 2, except the three SSI functions were weighted using %ncAA_{active} of the TK_{incorporated} and %F_{active}/%Y_{active} of the TK_{misincorporated} species, all derived from mass spectra. The methodology therefore normalised for variations in the active proportion of TK_{incorporated} and TK_{misincorporated} species, which varied from 11.5% to 28.3%, and hence provided even more clarity on the best-performing active variant subspecies. The only caveat remaining is that this analysis again averages out any specific effects of mixed dimer species. Analysis 3 therefore makes the most assumptions for deconvolution, but gave the best comparative analysis of TK_{incorporated} variant subspecies. The kinetic parameters for TK_{active} subspecies of each variant were also most amenable to interpretation through *in silico* molecular docking simulations, as discussed below.

Taking the discussed caveats into account, we were able to determine the kinetic parameters of (nc)AA-TK_{biocat}, (nc)AA-TK_{incorporated} and (nc)AA-TK_{active} species of each variant in order to determine, or at the very least compare, the relative catalytic performance of each variant. While, as expected, their absolute values varied significantly between TK_{biocat} and the TK_{active} subspecies, the relative changes in parameter trends were relatively constant (with the exception of 385pAMF) because the %TK_{misincorporated} and %TK_{high} remained comparable between variants. Analyses 2 and 3 revealed the ‘true’ catalytic potential of 385pAMF that was otherwise masked by low-activity TK_{misincorporated} and TK_{low}/TK_{inactive} subspecies. All kinetic parameters referred to henceforth correspond to only the TK_{active} subspecies of each variant (Table 2), unless stated otherwise.

Analysis of the kinetic parameters of (nc)AA-TK_{active}

While we expected a change in the catalytic performance upon tweaking the electronics of an important residue in the active site, it was difficult to predict what trend, if any, may emerge given the complexity of substrate binding and catalytic turnover.

The specific activity of 385pCNF was an impressive 43-fold greater than that of 385Y, while those of 385pAMF, 385pNTF and 385F increased 13-fold, 4-fold and 1.2-fold, respectively (Fig. 6A; Table 2). However, no clear trend was observed in the specific activity, K_m nor k_{cat} (Fig. 6B; Table 2) as a function of the aromatic ring electron density of residue 385, which confirmed that the altered active-site interactions were more complex than simply relating to ring electron density.

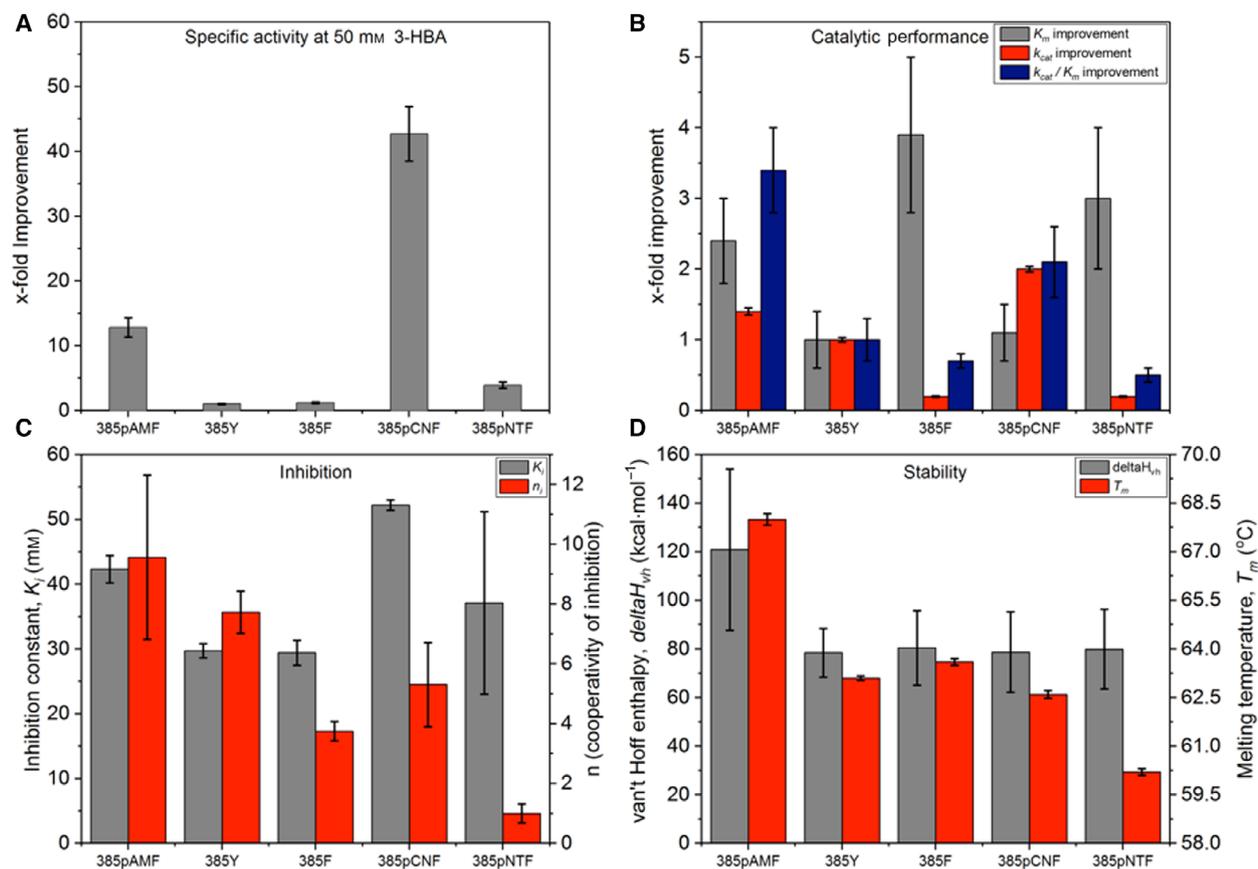


Fig. 6. Summary of the catalytic performance (A and B), acceptor substrate inhibition (C) and the stability (D) of the TK_{active} subspecies of variants 385pAMF, 385Y, 385F, 385pCNF and 385pNTF. The mutants are ranked in order of decreasing aromatic ring electron density from left to right. (A) Specific activities at 50 mM 3-HBA relative to mutant 385Y; (B) catalytic performance relative to variant 385Y, in terms of improvement of K_m (grey), k_{cat} (red) and k_{cat}/K_m (blue); (C) acceptor substrate inhibition profiles, in terms of K_i (grey, left y -axis) and n_i (red, right y -axis); and (D) thermal stability, in terms of T_m and van't Hoff enthalpy, ΔH_{vh} . Associated errors are the SEM fitting error for the SSI model function (see Materials and methods) or the thermal denaturation function [42–44]. $n = 3$.

Interestingly, the catalytic improvement of the two best-performing variants, 385pAMF and 385pCNF, relative to 385Y, was driven by two divergent evolutionary mechanisms. Incorporation of pAMF improved the catalytic efficiency, k_{cat}/K_m , by 240%, primarily via a 140% improvement in K_m (Table 2; Fig. 6B). Conversely, incorporation of pCNF at residue 385 instead improved k_{cat}/K_m by 110% almost exclusively via a 100% improvement in k_{cat} . The other two mutations, 385F and 385pNTF, resulted in large improvements in K_m (290% and 200% improvements, respectively), but even greater decreases in k_{cat} .

The theoretical total number of inhibitory substrate orientations and/or binding events, n_i , was correlated with the aromatic ring electron density of residue 385 (Fig. 6C; Table 2). 3-HBA could bind 385pAMF in as many as 10 inhibitory orientations, while that was

reduced to one in 385pNTF. Variant 385pCNF had a higher than expected n_i and was a slight anomaly to the general trend. There appeared to be no obvious trend in K_i as a function of aromatic ring electron density as variants 385Y, 385F and 385pNTF had similar K_i -values, while the introduction of pAMF and pCNF at residue 385 improved the K_i by 42% and 76%, respectively.

Overall, the substantially improved specific activity of pCNF at 50 mM 3-HBA relative to 385Y was explained by a combination of a 100% improvement in k_{cat} and substantially reduced substrate inhibition (through both a decrease in n_i and an increase in K_i).

Thermal stability of holo-transketolase variants

Previously, introduction of mutations that conferred rigidity into the cofactor-binding loop (e.g. H192P/

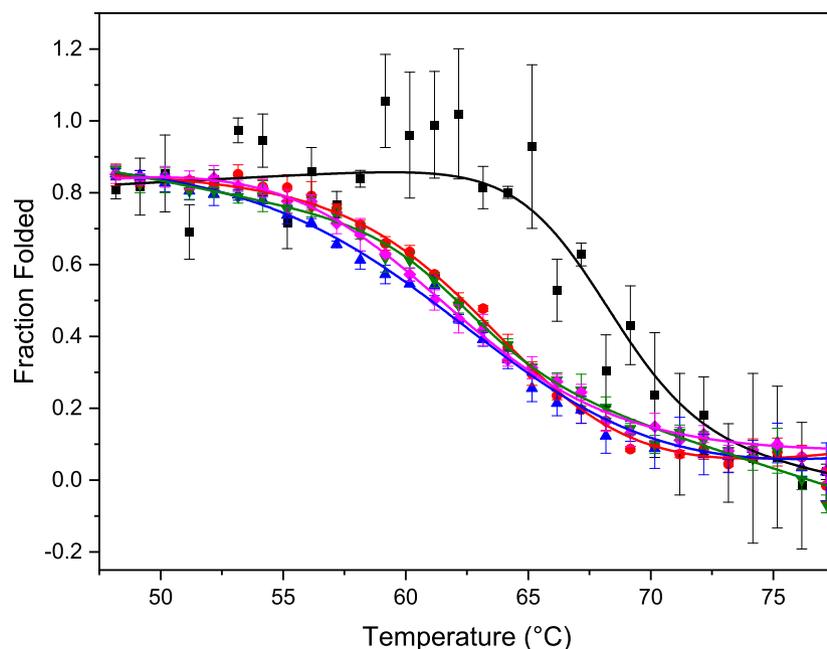


Fig. 7. The thermal unfolding of all TK subspecies (ncAA_{biocat} or AA_{biocat}) of variants 385pAMF (black), 385Y (red), 385F (blue), 385pCNF (green) and 385pNTF (magenta). The apparent thermal stability was determined by the change in the fluorescence emission ratio (350/330) as a function of time, and fitted to a two-state model of thermal denaturation, as described previously [42–44]. Error bars represent the standard error of the mean (SEM), $n = 3$.

H282P) led to improved thermal stability of wild-type transketolase by 5 °C [41]. We postulated that a decrease in aromatic ring electron density via electron-withdrawing substitutions may also strengthen π - π stacking interactions between residue 385 and active-site nonpolar aromatic residues, with the potential to increase active-site rigidity and hence thermal stability.

The methodology of Analyses 1 and 2 was reapplied to the experimental thermostability data of the five variants (Fig. 7) by similarly weighting a sum of two-state model functions for thermal denaturation [42–44] of each variant (Equation 4, Materials and Methods). In Analysis 2, the variants were globally fitted, with the stability parameters of 385Y and 385F shared, and the % (nc)AA-TK_{incorporated}, %385F-TK_{misincorporated} and %385Y-TK_{misincorporated} weighting parameters fixed for each variant. TK_{active} cannot currently be isolated from TK_{low}, and their relative stabilities are unknown; therefore, the stability of (nc)AA-TK_{active} cannot be calculated using the methodology of Analysis 3. The T_m values from Analyses 1 and 2 (Table 3) were within error for all variants, which indicated (nc)AA-TK_{biocat} stability was largely dominated by the most populous species, (nc)AA-TK_{incorporated}. The ΔH_{vh} values varied considerably more between Analyses 1 and 2, with the values from Analysis 2 generally higher than those of Analysis 1. This was potentially a consequence of the interactions between species and subspecies in heterodimers that have not been accounted for in the model. Since fewer caveats apply to Analysis 1, and Analysis 2 gleans little more

Table 3. Summary of the thermal stability of TK_{biocat} and TK_{incorporated} of wild-type, 385pAMF, 385Y, 385F, 385pCNF and 385pNTF. Parameters were determined by fitting the change in the fluorescence emission ratio (350/330) as a function of temperature to a two-state model of thermal denaturation [42–44], using the methodology of Analyses 1 and 2. Associated errors are the fitting error for the thermal denaturation function

Variant	Analysis 1 (TK _{biocat})		Analysis 2 (TK _{incorporated})	
	T_m (°C)	ΔH_{vh} (kcal/mol)	T_m (°C)	ΔH_{vh} (kcal/mol)
Wild-type	65.6 ± 0.6	77.2 ± 3.8	–	–
385pAMF	68.0 ± 1.2	120.9 ± 33.2	67.7 ± 0.6	244.1 ± 2.6
385Y	63.1 ± 0.5	78.4 ± 10.0	64.3 ± 0.6	72.8 ± 11.7
385F	63.6 ± 0.7	80.4 ± 15.3	63.3 ± 1.8	51.6 ± 13.8
385pCNF	62.6 ± 0.8	78.7 ± 16.5	62.1 ± 0.6	129.8 ± 28.6
385pNTF	60.2 ± 0.7	79.9 ± 16.4	60.7 ± 1.0	63.9 ± 15.6

information than Analysis 1, all stability parameters referred to henceforth correspond to those of only the TK_{biocat} species of each variant (Table 3), unless stated otherwise.

The T_m did not correlate linearly to any properties of the variants or mutation types, with the best observed for n_i of the variants ($R^2 = 0.73$). However, the thermal transition mid-point, T_m , increased nonlinearly with aromatic ring electron density (Fig. 6D), in the opposite direction to that predicted based on the strength of π - π stacking. The most stable variant, 385pAMF, was 7.8 °C more stable than the least

stable 385 p NTF variant, 4.9 °C higher than 385Y (3M), and remarkably, 2.4 °C higher than wild-type transketolase (Table 3). This was the first time that the aromatic substrate-accepting 3M variant had been engineered to be more stable than the wild-type [45].

The thermal denaturation of 385 p AMF had increased cooperativity, as indicated by the high value of ΔH_{vh} (Fig. 7; Table 3), which is itself indicative of tight overall packing and increased rigidity. The additional rigidity in 385 p AMF was not expected from the increased aromatic ring electron density, which would in theory form weaker hydrophobic interactions than the other four aromatic amino acids. Molecular modelling instead suggested that the shift in the *para*-substituted functional group, from an H-bond acceptor (p CNF and p NTF), to nonpolar (F), to an H-bond donor (p AMF and Y), increased the active-site and cofactor-loop rigidity. Analysis of the energy-minimised structures of 385 p AMF and 385Y showed the carbonyl backbone of G262 was within H-bonding distance of the $-NH_2$ and $-OH$ side chain groups, at 3.6 Å and 4.2 Å, respectively (Fig. 8A, C & D). The latter distance is large for an H-bond, and, if formed, such an interaction would be extremely weak. In both energy-minimised structures, a helix-turn was formed between residues D259-H261, as a result of

stabilisation by the interaction between residue 385 and G262. Neither an H-bond nor a helix-turn was observed in the other three less-stable variants (Fig. 8 B, E, F & G). G262 is located in the opposite subunit of the transketolase dimer and therefore provides both an anchor to increase the rigidity of the cofactor-binding loop, increasing ΔH_{vh} , and an additional intersubunit interaction, to increase T_m . The loss of secondary structure in at least three variants may explain why the ΔH_{vh} was much lower and roughly equivalent across these variants. Furthermore, the weakened or absent intersubunit H-bond, in addition to the introduction of bulkier, and polar, ring-substituted functional groups, may explain the gradual decrease in T_m observed from 385 p AMF to 385 p NTF.

Computational docking of 3-HBA into triple-mutant variants

The structure–function relationships driving the two divergent evolutionary mechanisms of p AMF and p CNF, and the trend in changing substrate inhibition across the variants were subsequently interrogated further. *In silico* molecular docking of 3-HBA into the active site of variant Y, and also into the computationally mutated, energy-minimised active sites of

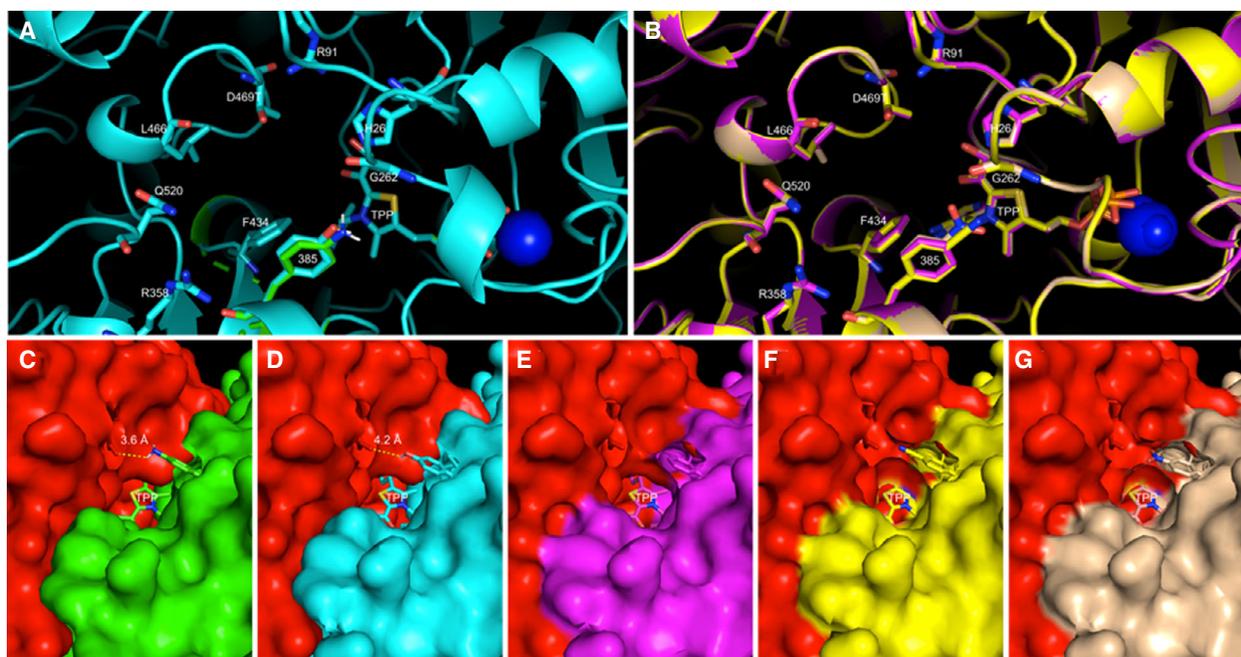


Fig. 8. The energy-minimised active-site structures of p AMF, Y (A; green and cyan), F, p CNF and p NTF (B; magenta, yellow and wheat). Important active-site residues are labelled in each. Note that the only significant structural change is the formation of a helix-turn between residues D259-H261 of subunit B. The possible H-bonds between residue 385 of subunit A and G262 of subunit B is shown for each variant in C) D) E) and F), respectively

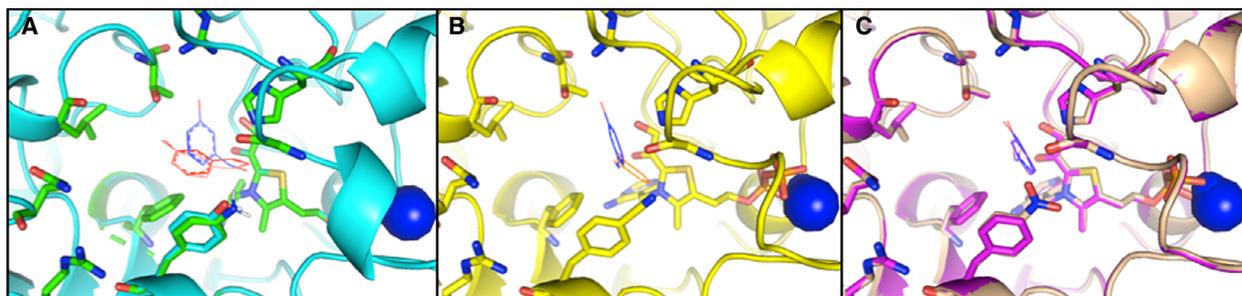


Fig. 9. Computational docking of 3-HBA into the active site of mutants *pAMF* (A: green), Y (A: cyan), F (C: magenta), *pCNF* (B: yellow) and *pNTF* (C: wheat). Only subcluster poses in catalytically productive orientations are shown. 3-HBA was bound in both Pocket A (red and salmon red) and Pocket B (blue and light blue) in mutants *pAMF* and Y, respectively (A); in two orientations in Pocket B of *pCNF* (B; blue and orange); and a single orientation in Pocket B of F and *pNTF* (C; purple-blue and blue).

385*pAMF*, 385F, 385*pCNF* and 385*pNTF*, provided insights into the molecular basis of the experimentally determined parameters of catalytic performance, acceptor substrate inhibition and protein stability. From these, we propose plausible structural explanations for the experimentally determined kinetic parameters for each variant.

For all analyses, the highest energy substrate subcluster, energy cluster 1 (EC1), shown in Fig. 9, was considered catalytically productive when the aldehyde moiety was oriented 0° or 180°, and catalytically inhibitive when oriented 90° or 270°, relative to the DHE-TPP intermediate. All subclusters of EC2 or lower were defined as inhibitory. Active-site interactions with the catalytically productive substrate poses were subsequently analysed to rationalise the experimentally determined catalytic data, in terms of: the number of productive, unproductive and inhibitory poses; binding pocket preference (% Pockets A and B); and predicted active-site interactions. The results are summarised in

Table 4. Summary of the number of catalytically productive/unproductive poses and pocket occupation of variants 385*pAMF*, 385Y, 385F, 385*pCNF* and 385*pNTF*, predicted by *in silico* molecular docking simulations

Variant	No. productive poses	No. unproductive poses	% Pocket A	% Pocket B
385 <i>pAMF</i>	2	3	48%	52%
385Y	2	3	18%	82%
385F	1	2	0%	100%
385 <i>pCNF</i>	2	1	0%	100% ^a
385 <i>pNTF</i>	1	1	0%	100%

^a385*pCNF* Pocket B was subdivided into Pocket B (46%) and the extended Pocket B* (54%).

Table 4, while the full results can be found in Supporting Information (Supporting Information Appendix, Supporting Information Text 3).

Comparison of variant Y and *pAMF*

In silico analysis predicted 3-HBA to bind to the active site of 385*pAMF* and 385Y in near-identical orientations (Fig. 9A), including two that were catalytically productive and three that were inhibitory. The high number of inhibitory subclusters in both 385*pAMF* and 385Y was consistent with the high cooperativity of inhibition, and high theoretical number of inhibitory binding orientations (n_i) observed experimentally. The two catalytically productive subclusters of each variant were split between the two binding pockets that had been previously identified [15] (Fig. 9A). The decreased aromatic ring electron density in 385Y relative to 385*pAMF* shifted the binding pocket preference from Pocket A to Pocket B, with 52% of productive poses in Pocket B of 385*pAMF* compared to 82% in 385Y. The observed shift to Pocket B may have been due to the loss of two interactions between 3-HBA and Pocket A of 385Y compared to that of 385*pAMF* (Fig. S4). The additional active-site interactions in 385*pAMF* were consistent with the decreased K_m , observed experimentally.

Analysis of 385*pCNF*

3-HBA was predicted to bind 385*pCNF* in two catalytically productive and one inhibitory subcluster of poses that were unique to 385*pCNF* (Fig. 9B) and hence could not be directly compared to the sets of subclusters that were common between variants 385*pAMF*/385Y and 385F/385*pNTF* (Fig. 6A). The

low number of *in silico* inhibitory poses agreed with the relatively low n_i determined experimentally. The presence of two catalytically productive subclusters likely contributed to the higher n_i value relative to 385F and 385 p NTF, and may also have contributed to the low substrate affinity as a result of a less-favourable enthalpic contribution.

The further decrease in aromatic ring electron density completely shifted the binding pocket preference from Pocket A to Pocket B, with 54% of productive poses occurring in a newly created extended region of Pocket B (Pocket B*). The catalytic subcluster in Pocket B* was predicted to bind weakly with few interactions between the active-site and 3-HBA, and an unfavourable interaction between the polar aldehyde moiety of 3-HBA and the nonpolar ring of F434 (Fig. S5). Nevertheless, the aldehyde moiety was positioned very close to the DHE-TPP intermediate in the extended pocket and had a Bürgi–Dunitz angle (the geometric angle between the carbonyl plane and the nucleophile-carbon line) closest to the theoretically best Bürgi–Dunitz angle of (124° versus 107°, respectively). The prediction of one low-affinity, highly productive and one high-affinity, lowly productive binding pocket within the active site of 385 p CNF is in agreement with the high K_m , k_{cat} and higher than expected n_i determined experimentally.

Comparison of 385F and 385 p NTF

The single catalytically active subcluster of 385F and 385 p NTF was predicted to bind exclusively in Pocket B in identical orientations (Fig. 9C), while each variant had two and one additionally inhibitory subclusters, respectively. The inhibitory subclusters were low in energy and number, again supporting the low number of inhibitory substrate orientations observed experimentally. Furthermore, the higher number of inhibitory subclusters and a higher overall proportion of inhibitory poses (20%) also explains the increased substrate inhibition observed in 385F compared to 385 p NTF (0.5% of poses were inhibitory).

385F and 385 p NTF were predicted to have near-identical interactions with 3-HBA in their catalytically productive orientations, the only difference being a stronger π – π stacking interaction with residue F385 compared to only weak Van der Waal's forces with residue p NTF385 (Fig. S6). This difference was not observed in the experimentally determined k_{cat} and K_m , however, highlighting the caution that must be taken in the interpretation of computational docking, which at best provides only guidance for possible mechanistic explanations, and to inform future

experimental work. Finally, the aldehyde moiety of 3-HBA was in close proximity to H473, potentially explaining the low K_m values for variants 385F and 385 p NTF.

Discussion

A comparison of the kinetic and stability parameters of each variant before and after accounting for misincorporation and the oxidised TK subspecies demonstrated the impact that this can have on the kinetic parameters derived directly from the experimental data. Previously, studies into the activity and stability of ncAA-incorporated variants have confirmed, but often not quantified, incorporation and misincorporation by mass spectrometry. We propose that determination of the fidelity of incorporation should be a standard procedure in studies similar to ours to allow a more accurate comparison of the catalytic performances of two or more ncAA-incorporated variants.

Through evolution and expression optimisation of tRNA/synthetase pairs, the fidelity of ncAA incorporation has improved greatly over the years. It is relatively unsurprising that the ncAA-incorporation fidelities obtained for GFP are often high, given that most tRNA/synthetase pairs have been evolved using GFP as the target protein. However, incorporation efficiency and fidelity can be protein-/sequence-dependent because of a range of factors, such as mRNA secondary structure, promoter sequence, rate of translation, protein length, amino acid sequence and the interaction between sequential amino acids and their cognate tRNA/synthetase pairs as the nascent polypeptide chain is synthesised. Differentiating between structurally similar natural and noncanonical amino acids is often challenging, even for evolved orthogonal tRNA/synthetase pairs. In these cases, higher levels of misincorporation are unavoidable and must be accounted for in kinetic comparisons.

This study began with a highly evolved variant that had already been optimised for acceptance of aromatic aldehyde substrates through saturation mutagenesis with natural amino acids at residue 385 [15]. However, the improved function had led to a trade-off in stability, which had been recovered through the engineering of molecular dynamics in a previous study, but never improved beyond the stability of wild-type TK [45]. The expansion of saturation mutagenesis to include ncAAs has now led to observed improvements in both catalytic activity and stability simultaneously, which demonstrates the benefits of their inclusion in designer library approaches. Even a small library of variants at a single active-site residue generated a broad range of

catalytic and stabilising properties. Variant 385pAMF had the highest thermal stability, a relatively high maximal activity and strong substrate binding, but high susceptibility to substrate inhibition. Variant 385pCNF gave high catalytic turnover and lower susceptibility to substrate inhibition, but weaker substrate affinity. Variants 385F and 385pNTF had lower maximum activities, but strong substrate binding and low susceptibility to substrate inhibition.

To our knowledge, S385pAMF/D469T/R520Q is the first example in which both catalytic activity and stability are simultaneously improved via site-specific ncAA incorporation into an enzyme active site, and demonstrates the benefits of both fine-tuning preevolved residues using ncAAs, but also expanding directed evolution or designer libraries to include ncAAs in general.

Materials and methods

Chemicals and reagents

Tris/HCl was purchased from VWR International (Lutterworth, UK), and guanidine-HCl was purchased from Life Technologies Ltd. HPA was synthesised by reacting bromopyruvic acid with LiOH, as described previously [46]. *Para*-cyanophenylalanine was purchased from Bachem (California, USA). All other chemical reagents were purchased from Sigma-Aldrich (Poole, UK).

Mutagenesis at residue 385

Mutagenesis was carried out using QuikChange site-directed mutagenesis using the manufacturer's protocol (Stratagene, Cambridge, UK) and the following primer and its reverse complement, designed for specific mutations at residue 385:

5'-GCTGACCTGGCGCCGTAGACCTGACCCTGTGG-3'
ncAA mutants

5'-GCTGACCTGGCGCCGTTTACCTGACCCTGTGG-3'
Mutant F

5'-GCTGACCTGGCGCCGCAGACCTGACCCTGTGG-3'
Mutant Q

The *dpnI*-digested PCR product was transformed into XL10-gold competent cells and the plasmid subsequently isolated using a Qiagen Miniprep Kit (Stratagene, Cambridge, UK).

Preparation of and cotransformation into competent C321.ΔA.exp 'amberless' cells

The 'amberless' *E. coli* strain, *C321.ΔA.exp*, a gift from George Church (Addgene plasmid #49018), was used as the expression strain for all variants. A 50 mL culture of

C321.ΔA.exp in LB was grown to an $OD_{600} \approx 0.5$ in a 250-mL shake flask and subsequently transferred to two prechilled 50-mL falcon tube and cooled for 10 min on ice. All consumables required were cooled on ice for the duration of the procedure. The cells were centrifuged at 2700 *g* for 10 min at 4 °C, the supernatant discarded and the pellet resuspended in 1.6 mL precooled 100 mM CaCl₂ for 30 min on ice. Centrifugation followed by resuspension and incubation was repeated for each falcon tube. Finally, the cells were combined into a single tube and 0.5 mL prechilled 80% glycerol was added. The resulting chemically competent cells were frozen in liquid nitrogen and stored at -80 °C until required. The chemically competent *C321.ΔA.exp* cells were cotransformed with *pUltra*, a gift from Peter Schultz (Addgene plasmid #48215), encoding the ncAA-incorporation system [20], and pQR791, encoding a constitutively expressed transketolase variant.

Enzyme preparation and enzyme kinetics

Variants of the transketolase mutant S385Y/D469T/R520Q were coexpressed with the ncAA-incorporation machinery from the *pUltra* plasmid for eight hours in *C321.ΔA.exp* cells in the presence of 1 mM ncAA and 1 mM IPTG. The resulting cell pellet was lysed and purified as described previously [47]. Purified transketolase was ultrafiltered four times using Amicon Ultra-4 10k MWCO centrifugal filter to remove excess imidazole and cofactors. Protein concentration was determined by absorbance at 280 nm in 6 M guanidine-HCl and 20 mM sodium phosphate, pH 6.5. Absorbance was measured using a NanoDrop spectrophotometer; the molecular weight of each variant was based on the wild-type monomeric molecular weight of 73035.5 g·mol⁻¹ and an extinction coefficient (ϵ) of 92630 L·mol⁻¹·cm⁻¹, modified for each variant.

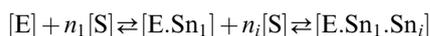
Kinetic parameters were obtained at saturating 50 mM HPA and 3–50 mM 3-HBA. About 80 μ L 0.6–1.0 mg·mL⁻¹ TK was incubated with 20 μ L of 10 × cofactor solution (24 mM ThDP, 90 mM MgCl₂) for 30 min, and the reaction initiated with 100 μ L 2x 3-HBA in 50 mM Tris/HCl, pH 7.0. All reactions were carried out in triplicate in glass vials at 22 °C. Samples of each reaction were quenched every 30 min for 180 min by addition of 380 μ L 0.1% TFA to 20 μ L sample and centrifuged at 13 000 rpm for 3 min, and the supernatant analysed by HPLC, as described previously [14]. The TK concentration in each reaction was between 0.07 and 0.3 mg·mL⁻¹. Higher TK concentrations were used for the 3-HBA reaction due to the slower conversion. All data were fitted by nonlinear regression to either the sequential noncompetitive inhibition (NCI; Equation 1) or sliding-substrate inhibition (SSI; Equation 2) model to determine the kinetic parameters of each variant.

$$v = \frac{V_{\max}}{\frac{K_m}{[S]} + 1 + \frac{[S]}{K_i}} \quad (1)$$

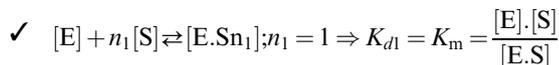
$$v = \frac{V_{\max}}{\frac{K_m}{[S]} + 1 + \frac{[S]^{n-1}}{K_i^{n-1}}} \quad (2)$$

Derivation of the modified Michaelis–Menten equation

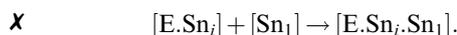
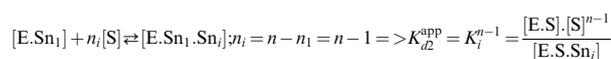
The modified Michaelis–Menten function was derived in a similar way to the standard Michaelis–Menten function, except using the following chemical equilibria, left, that describe a single catalytically productive enzyme–substrate binding event, $n_1 = 1$, that is completely inhibited at high [S] by multiple (n_i) inhibitory enzyme–substrate interactions. The function assumes that the catalytically productive binding event occurs prior to the inhibitory binding event in a sequential reaction order. The chemical equilibria marked with a cross are those which become obsolete once the reaction is defined as kinetically sequential. The respective dissociation constants, right, are derived using the law of mass action.



↓



✓



where [E] is the enzyme concentration, [S] is the substrate concentration, n_1 and n_i are the theoretical number of molecules, or number of orientations, of substrate binding in the first (catalytically productive) and second (inhibitory) binding event, n is the theoretical total number, or total orientations, of productive and inhibitory substrate molecules that can bind the enzyme and K_{d1} and K_{d2}^{app} are the apparent dissociation constants of the first catalytically productive binding event and all subsequent inhibitory binding events, respectively.

The inhibitory binding events are defined as either multiple substrate molecules that simultaneously bind to the enzyme active site, or as a single substrate molecule that interacts with the enzyme active site in multiple inhibitory binding poses. The n_i number of inhibitory poses are assumed to have slightly different binding inhibition

constants, K_i , which are represented by a single inhibitory dissociation constant, K_{d2}^{app} . In other words, the individual inhibitory dissociation constants for each binding pose are convoluted or 'averaged' by multiplying and dividing the binding constants together to give a single representative inhibitory binding constant, such that.

$$K_{d1} = K_{d1}^n = K_{d1}^1 = K_m; \text{ and}$$

$$K_{d2}^{\text{app}} = K_{i1} \cdot K_{i2} \cdot K_{i3} \cdot K_{i4} \cdot \dots \cdot K_{ini} = K_i^{n-1} = K_i^{n-1}$$

In theory, the substrate can 'slide' between inhibitory binding poses. 'Sliding' is significantly more likely to be observed in catalytic reactions that are dominated by hydrophobic enzyme–substrate interactions, which result in much less-well defined enzyme–substrate orientations compared to electrostatic interactions.

The total enzyme concentration, $[E_T]$, can be expressed as the following:

$$[E_T] = [E] + [E.S] + [E.S.Sn_i].$$

Substitution of [E.S] and [E.S.Sn_i] for their dissociation constant expressions (above right) gives:

$$[E_T] = [E] + \frac{[E].[S]}{K_m} + \frac{[E.S].[S]^{n-1}}{K_i^{n-1}} = [E] + \frac{[E].[S]}{K_m} + \frac{[E].[S].[S]^{n-1}}{K_m.K_i^{n-1}}$$

which rearranges and simplifies to the following:

$$[E] = \frac{[E_T]}{1 + \frac{[S]}{K_m} + \frac{[S].[S]^{n-1}}{K_i^{n-1}}} = \frac{[E_T]}{1 + \frac{[S]}{K_m} + \frac{[S]^n}{K_m.K_i^{n-1}}}$$

Initial reaction velocity can be derived by substitution of

$$K_m = \frac{[E].[S]}{[E.S]} \text{ into } v_p = k_{\text{cat}}.[E.S]:$$

$$v_p = k_{\text{cat}}.[E.S] = \frac{k_{\text{cat}}.[E].[S]}{K_m} = \frac{k_{\text{cat}}.[E_T].[S]}{K_m \left(1 + \frac{[S]}{K_m} + \frac{[S]^n}{K_m.K_i^{n-1}} \right)} = \frac{k_{\text{cat}}.[E_T]}{\frac{K_m}{[S]} + 1 + \frac{[S]^{n-1}}{K_i^{n-1}}}$$

Finally, substitution of $v_{\max} = k_{\text{cat}}.[E_T]$ into the above equation gives the SSI model:

$$v = \frac{V_{\max}}{\frac{K_m}{[S]} + 1 + \frac{[S]^{n-1}}{K_i^{n-1}}}$$

The SSI model above contrasts to the noncompetitive inhibition (NCI) model (Equation 1), which describes inhibition by a single substrate molecule in a single orientation. In this case, the SSI model simplifies to the NCI model when $n = 2$, and therefore $n_i = 1$.

Determination of kinetic parameters of (nc)AA-TK_{biocat}, (nc)AA-TK_{incorporated} and (nc)AA-TK_{active}

The analysed mass spectrometry data were combined with the experimentally determined activity data and then

globally fitted to the weighted sum of three SSI model functions (Equation 3).

$$v = \% A * \frac{V_{\max}(A) \cdot [S]}{K_m(A) + [S] + \frac{[S] \cdot [S]^{n(A)-1}}{K_i(A)^{n(A)-1}}} + \% B * \frac{V_{\max}(B) \cdot [S]}{K_m(B) + [S] + \frac{[S] \cdot [S]^{n(B)-1}}{K_i(B)^{n(B)-1}}} + \% C * \frac{V_{\max}(C) \cdot [S]}{K_m(C) + [S] + \frac{[S] \cdot [S]^{n(C)-1}}{K_i(C)^{n(C)-1}}} \quad (3)$$

where A, B and C are %ncAA-TK_{incorporated}, %F-TK_{misincorporated} and %Y-TK_{misincorporated} (Analysis 2) or %ncAA-TK_{active}, %TK_{active} of F-TK_{misincorporated} and %TK_{active} of Y-TK_{misincorporated} (Analysis 3).

Mass spectrometry

LC-ESI-MS was performed using an Agilent 1100/1200 LC system connected to a 6510A QTOF mass spectrometer (Agilent, UK). Samples of 10 μL TK at 0.2 $\mu\text{g} \cdot \mu\text{L}^{-1}$ were injected onto an Agilent PLRP-S (150 mm \times 2.1 mm, 1000 \AA , 8 μm) column, maintained at 30 $^{\circ}\text{C}$. Two mobile phases A (5% MeCN in aqueous 0.1% formic acid) and B (95% MeCN, 5% water, 0.1% formic acid) were used at 0.3 $\text{mL} \cdot \text{min}^{-1}$. The column was pre-equilibrated at 25% B for 1.9 min, before injection, held for 1 min further at 25% B, and then a gradient elution increased B to 99% over 16 min. After 2 min, B was decreased to 25% over 0.1 min. The QTOF mass spectrometer scanned m/z from 100 to 3100 Da. Positive electrospray ionisation (ESI) was used with 4000 V capillary voltage, fragmentor at 175 V, skimmer at 65 V and octopole RF peak at 750 V. Nitrogen was used as the nebuliser and desolvation gas at a flow of 5 $\text{L} \cdot \text{min}^{-1}$. Spectra were acquired every second with an acquisition time of 1000 msec/spectrum. Lockspray was used during analysis to maintain mass accuracy. Data were processed in MassHunter software (version B.07.00) and deconvolved using the maximum entropy deconvolution algorithm.

SDS/PAGE analysis

The total concentration of prepared lysate was determined by a Bradford assay and diluted to 1 $\text{mg} \cdot \text{mL}^{-1}$. 25 μL 2x Laemmli buffer (Bio-Rad Laboratories, Watford, UK) was added to 25 μL sample and heated for 10 min at 95 $^{\circ}\text{C}$ to denature the protein. 20 μL sample was then loaded into each well and analysed by an Amersham Imager 600.

Thermal denaturation measurements

The T_m values of TK variants were measured in the Unit (Unchained Laboratories, Wetherby, UK) via their intrinsic fluorescence emission ratio (350/330 nm). The microcuvette arrays were loaded with 9 μL of 0.5 $\text{mg} \cdot \text{mL}^{-1}$ sample and excited with a 266 nm laser. The fluorescence was measured as a function of temperature in the range of 30–90 $^{\circ}\text{C}$ with steps of 1 $^{\circ}\text{C}$, equilibration time of 30 s at each temperature. Thermal denaturation curves were analysed by fitting the baseline and single transition to a two-state model [42–44].

The analysed mass spectrometry data were combined with the experimentally determined stability data and then globally fitted to the weighted sum of three SSI model functions (Equation 4).

$$y = \% A * f_a + \% B * f_b + \% C * f_c \quad (4)$$

where A, B and C are %ncAA-TK_{incorporated}, %F-TK_{misincorporated} and %Y-TK_{misincorporated} (Analysis 2), and f is the two-state model function [42–44]:

$$f = \frac{(I_N + aT) + (I_D + bT) \cdot \exp\left[\frac{\Delta H_{vh}}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T}\right)\right]}{1 + \exp\left[\frac{\Delta H_{vh}}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T}\right)\right]}$$

where I_N and I_D are the native and denatured state signals, respectively, a and b are the slopes of the native and denatured state signal baselines, respectively, R is the molar gas constant and ΔH_{vh} is the van't Hoff enthalpy at the thermal transition mid-point, T_m .

Computational docking

The crystal structure of S385Y/D469T/R520Q was mutated at position 385 to $p\text{AMF}$, Y, F, $p\text{CNF}$ and $p\text{NTF}$ using the Swiss-Sidechain-plugin in PyMol Molecular Graphics System (Schrödinger, USA), and energy-minimised using Charm forcefield Adopted Basic NR, Implicit Generalised Born solvent model, True SHAK E constant and 1000 steps in Discovery Studio 2.0 (Accelrys, Inc. San Diego, CA, USA). Each variant, with the enamine-ThDP intermediate present, was subsequently stripped of all crystallographic waters and docked in Autodock 4.2.6. Ligands were obtained as mol2 files and assigned three-dimensional coordinates. The explorable space for docking was defined as previously [15]. For each search, a Lamarckian genetic algorithm was run 200 times with a maximum number of 25 million energy evaluations. Resulting poses were analysed and checked for hydrogen bonding in PyMol and Discovery Studios 2017. Poses were first split into clusters that differed in energy by 0.04 ($\text{kcal} \cdot \text{mol}^{-1}$), and designated energy clusters (e.g. EC1). These energy clusters were subdivided into subclusters, defined as catalytically productive when the aldehyde moiety was oriented 0 $^{\circ}$ or 180 $^{\circ}$, and

catalytically inhibitive when oriented 90° or 270°, relative to the DHE-TPP intermediate. All subclusters of EC2 or lower were defined as inhibitory. For molecular analysis, interactions facing 180° from the DHE-TPP intermediate were rotated 180° about the centre and plane of the substrate aromatic ring.

Acknowledgements

We thank Dr. Kersti Karu for mass spectrometric measurements in the UCL Department of Chemistry Mass Spectrometry Facility. We also thank Prof Peter Schultz, Prof George Church and Prof John Ward for donations of the *E. coli* strain *C321.ΔA.exp* (Addgene plasmid #49018), the plasmid *pUltra* (Addgene plasmid #48215) and the transketolase plasmid *pQR791*, respectively. Finally, we thank the Engineering and Physical Sciences Research Council (EPSRC) for funding (EP/L015218/1) via the Centre for Doctoral Training in Emergent Macromolecular Therapies.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

HCW carried out all experimental procedures, unless stated otherwise. PAD and HCW conceived and designed the experiments, wrote the main manuscript text and HCW prepared all figures. All authors reviewed the manuscript.

References

- 1 Wohlgemuth R (2009) C2-Ketol elongation by transketolase-catalyzed asymmetric synthesis. *J Mol Catal B Enzym* **61**, 23–29.
- 2 Kobori Y, Myles DC & Whitesides GM (1992) Substrate specificity and carbohydrate synthesis using transketolase. *J Org Chem* **57**, 5899–5907.
- 3 Hailes HC, Dalby PA, Lye GJ & Ward JM (2009) Biocatalytic approaches to ketodiols and aminodiols. *Chim Oggi* **27**, 28–31.
- 4 Hecquet L, Bolte J & Demuyne C (1996) Enzymatic synthesis of ‘natural-labeled’ 6-deoxy-L-sorbose precursor of an important food flavor. *Tetrahedron* **52**, 8223–8232.
- 5 Egan RM & Sable HZ (1981) Transketolase kinetics. The slow reconstitution of the holoenzyme is due to rate-limiting dimerization of the subunits. *J Biol Chem* **256**, 4877–4883.
- 6 Lindqvist Y, Schneider G, Ermler U & Sundstrom M (1992) Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 Å resolution. *EMBO J* **11**, 2373–2379.
- 7 Littlechild J, Turner N, Hobbs G, Lilly M, Rawas A & Watson H (1995) Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta Crystallogr Sect D Biol Crystallogr* **51**, 1074–1076.
- 8 Wilkinson HC & Dalby PA (2019) Novel insights into transketolase activation by cofactor binding identifies two native species subpopulations. *Sci Rep* **9**, 16116.
- 9 Hibbert EG, Senussi T, Costelloe SJ, Lei W, Smith MEB, Ward JM, Hailes HC & Dalby PA (2007) Directed evolution of transketolase activity on non-phosphorylated substrates. *J Biotechnol* **131**, 425–432.
- 10 Hibbert EG, Senussi T, Smith MEB, Costelloe SJ, Ward JM, Hailes HC & Dalby PA (2008) Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J Biotechnol* **134**, 240–245.
- 11 Cázares A, Galman JL, Crago LG, Smith MEB, Strafford J, Ríos-Solís L, Lye GJ, Dalby PA & Hailes HC (2010) Non- α -hydroxylated aldehydes with evolved transketolase enzymes. *Org Biomol Chem* **8**, 1301–1309.
- 12 Galman JL, Steadman D, Bacon S, Morris P, Smith MEB, Ward JM, Dalby PA & Hailes HC (2010) α , α' -Dihydroxyketone formation using aromatic and heteroaromatic aldehydes with evolved transketolase enzymes. *Chem Commun* **46**, 7608–7610.
- 13 Payongsri P, Steadman D, Strafford J, MacMurray A, Hailes HC & Dalby PA (2012) Rational substrate and enzyme engineering of transketolase for aromatics. *Org Biomol Chem* **10**, 9021–9029.
- 14 Payongsri P, Steadman D, Hailes HC & Dalby PA (2015) Second generation engineering of transketolase for polar aromatic aldehyde substrates. *Enzyme Microb Technol* **71**, 45–52.
- 15 Affaticati PE, Dai S-B, Payongsri P, Hailes HC, Tittmann K & Dalby PA (2016) Structural analysis of an evolved transketolase reveals divergent binding modes. *Sci Rep* **6**, 35716.
- 16 Strafford J, Payongsri P, Hibbert EG, Morris P, Bath SS, Steadman D, Smith MEB, Ward JM, Hailes HC & Dalby PA (2012) Directed evolution to re-adapt a co-evolved network within an enzyme. *J Biotechnol* **157**, 237–245.
- 17 Yu H & Dalby PA (2018) Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics. *Proc Natl Acad Sci USA* **115**, E11043–E11052.
- 18 Drienovská I, Mayer C, Dulson C & Roelfs G (2018) A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat Chem* **10**, 946–952.
- 19 Li JC, Liu T, Wang Y, Mehta AP & Schultz PG (2018) Enhancing protein stability with genetically encoded noncanonical amino acids. *J Am Chem Soc* **140**, 15997–16000.

- 20 Chatterjee A, Sun SB, Furman JL, Xiao H & Schultz PG (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **52**, 1828–1837.
- 21 Young TS, Ahmad I, Yin JA & Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J Mol Biol* **395**, 361–374.
- 22 Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G & Schultz PG (2011) An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894–1900.
- 23 Cooley RB, Karplus PA & Mehl RA (2014) Gleaning unexpected fruits from hard-won synthetases: probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBioChem* **15**, 1810–1819.
- 24 Cooley RB, Feldman JL, Driggers CM, Bundy TA, Stokes AL, Karplus PA & Mehl RA (2014) Structural basis of improved second-generation 3-nitro-tyrosine tRNA synthetases. *Biochemistry* **53**, 1916–1924.
- 25 Carlsson ACC, Scholfield MR, Rowe RK, Ford MC, Alexander AT, Mehl RA & Ho PS (2018) Increasing enzyme stability and activity through hydrogen bond-enhanced halogen bonds. *Biochemistry* **57**, 4135–4147.
- 26 Ugwumba IN, Ozawa K, Xu ZQ, Ely F, Foo JL, Herlt AJ, Coppin C, Brown S, Taylor MC, Ollis DL *et al.* (2011) Improving a natural enzyme activity through incorporation of unnatural amino acids. *J Am Chem Soc* **133**, 326–333.
- 27 Yu Y, Lv X, Li J, Zhou Q, Cui C, Hosseinzadeh P, Mukherjee A, Nilges MJ, Wang J & Lu Y (2015) Defining the role of tyrosine and rational tuning of oxidase activity by genetic incorporation of unnatural tyrosine analogs. *J Am Chem Soc* **137**, 4594–4597.
- 28 Burke AJ, Lovelock SL, Frese A, Crawshaw R, Ortmyer M, Dunstan M, Levy C & Green AP (2019) Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **1**, 219–223.
- 29 Ring M, Armitage IM & Huber RE (1985) m-Fluorotyrosine substitution in β -galactosidase; Evidence for the existence of a catalytically active tyrosine. *Biochem Biophys Res Commun* **131**, 675–680.
- 30 Parsons JF, Xiao G, Gilliland GL & Armstrong RN (1998) Enzymes harboring unnatural amino acids: Mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* **37**, 6286–6294.
- 31 Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF & Tirrell DA (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. *Angew Chemie – Int Ed* **40**, 1494–1496.
- 32 Votchitseva YA, Efremenko EN & Varfolomeyev SD (2006) Insertion of an unnatural amino acid into the protein structure: preparation and properties of 3-fluorotyrosine-containing organophosphate hydrolase. *Russ Chem Bull* **55**, 369–374.
- 33 Budisa N, Wenger W & Wiltschi B (2010) Residue-specific global fluorination of *Candida antarctica* lipase B in *Pichia pastoris*. *Mol Biosyst* **6**, 1630–1639.
- 34 Baker PJ & Montclare JK (2011) Enhanced refoldability and thermoactivity of fluorinated phosphotriesterase. *ChemBioChem* **12**, 1845–1848.
- 35 Jackson JC, Duffy SP, Hess KR & Mehl RA (2006) Improving nature's enzyme active site with genetically encoded unnatural amino acids. *J Am Chem Soc* **128**, 11124–11127.
- 36 Gung BW, Xue X & Reich HJ (2005) The strength of parallel-displaced arene-arene interactions in chloroform. *J Org Chem* **70**, 3641–3644.
- 37 Wang L & Schultz PG (2001) A general approach for the generation of orthogonal tRNAs. *Chem Biol* **8**, 883–890.
- 38 Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA *et al.* (2013) Genomically recoded organisms expand biological functions. *Science (80-.)* **342**, 357–360.
- 39 Zheng Y, Lajoie MJ, Italia JS, Chin MA, Church GM & Chatterjee A (2016) Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. *Mol Biosyst* **12**, 1746–1749.
- 40 Wilkinson HC & Dalby PA (2020) The Two-Species Model of transketolase explains donor substrate-binding, inhibition and heat-activation. *Sci Rep* **10**, 4148.
- 41 Yu H, Yan Y, Zhang C & Dalby PA (2017) Two strategies to engineer flexible loops for improved enzyme thermostability. *Sci Rep* **7**, 41212.
- 42 Santoro MM & Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl a-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068.
- 43 Zhang C, Samad M, Yu H, Chakroun N, Hilton D & Dalby PA (2018) Computational design to reduce conformational flexibility and aggregation rates of an antibody Fab fragment. *Mol Pharm* **15**, 3079–3092.
- 44 Robinson MJ, Matejtschuk P, Longstaff C & Dalby PA (2019) Selective stabilization and destabilization of protein domains in tissue-type plasminogen activator using formulation excipients. *Mol Pharm* **16**, 744–755.
- 45 Yu H & Dalby PA (2018) Exploiting correlated molecular-dynamics networks to counteract enzyme activity–stability trade-off. *Proc Natl Acad Sci USA* **115**, E12192–E12200.
- 46 Morris KG, Smith MEB, Turner NJ, Lilly MD, Mitra RK & Woodley JM (1996) Transketolase from *Escherichia coli*: a practical procedure for using the biocatalyst for asymmetric carbon-carbon bond synthesis. *Tetrahedron Asym* **7**, 2185–2188.

47 Martinez-Torres RJ, Aucamp JP, George R & Dalby PA (2007) Structural stability of *E. coli* transketolase to urea denaturation. *Enzyme Microb Technol* **41**, 653–662.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary text. Full analysis of ncAA-incorporation.

Supplementary equation S1. The random-order SSI model.

Fig. S1. The mass spectra of A) 385*p*AMF, B) 385Y C) 385F and D) 385*p*NTF.

Fig. S2. SDS-PAGE of 1 mg/mL lysate of A) 385*p*CNF and 385*p*NTF, B) 385*p*AMF, 385Y and 385F; and C) 0.2 mg/mL of purified 385*p*CNF.

Fig. S3. A fit of the experimental kinetic data of variants 385*p*AMF (black), 385Y (red), 385F (blue), 385*p*CNF (green) and 385*p*NTF (magenta) towards 3-HBA to the randomly-ordered SSI model.

Fig. S4. The predicted enzyme-substrate interactions of Pocket A (A and C) and Pocket B (B and D) of variants *p*AMF and Y, respectively.

Fig. S5. The predicted enzyme-substrate interactions of Pocket B (A) and the extended region of pocket B, B*, (B) of variant *p*CNF.

Fig. S6. The predicted Pocket B enzyme-substrate interactions of variants F and *p*NTF, respectively.

Table S1. The random-order SSI model.

Table S2. The expected (Q_{exp} , F_{exp} , Y_{exp} and ncAA_{exp}) and observed (Q_{obs} , F_{obs} , Y_{obs} and ncAA_{obs}) peak centres for variants 385*p*AMF, 385Y, 385F, 385*p*CNF and 385*p*NTF.