

# Engineering a Seven Enzyme Biotransformation using Mathematical Modelling and Characterized Enzyme Parts

William Finnigan,<sup>[a]</sup> Rhys Cutlan,<sup>[b]</sup> Radka Snajdrova,<sup>[c]</sup> Joseph P. Adams,<sup>[c]</sup>  
Jennifer A. Littlechild,<sup>[a]</sup> and Nicholas J. Harmer<sup>\*[a, b]</sup>

Multi-step enzyme reactions offer considerable cost and productivity benefits. Process models offer a route to understanding the complexity of these reactions, and allow for their optimization. Despite the increasing prevalence of multi-step biotransformations, there are few examples of process models for enzyme reactions. From a toolbox of characterized enzyme parts, we demonstrate the construction of a process model for a seven enzyme, three step biotransformation using isolated enzymes. Enzymes for cofactor regeneration were employed to

make this *in vitro* reaction economical. Good modelling practice was critical in evaluating the impact of approximations and experimental error. We show that the use and validation of process models was instrumental in realizing and removing process bottlenecks, identifying divergent behavior, and for the optimization of the entire reaction using a genetic algorithm. We validated the optimized reaction to demonstrate that complex multi-step reactions with cofactor recycling involving at least seven enzymes can be reliably modelled and optimized.

## Introduction

Biocatalysis, the use of isolated enzymes or whole cells to perform chemical reactions, is increasingly the route of choice in the chemical and particularly the pharmaceutical industries.<sup>[1]</sup> There is increasing interest in developing one-pot reactions, where multiple catalysts work together to complete chemical pathways. Multi-step pathways offer excellent cost and productivity benefits,<sup>[2]</sup> driving reversible processes to completion without the need to isolate substrates or products at each step.<sup>[3]</sup> Enzymes are well suited to these reactions, operating in similar aqueous conditions at low temperatures and pressures. In contrast, the use of several chemical steps together will often have differing operational requirements.<sup>[4]</sup> Using isolated enzymes in biotransformations offers advantages<sup>[5]</sup> including reduced complexity and cost in downstream processing,<sup>[6]</sup> fewer side reactions<sup>[7]</sup> and the removal of rate limiting diffusion across cellular membranes.<sup>[8]</sup> Most attractively, parameters such as enzyme or substrate concentration, co-solvents, pH or temperature, can easily be manipulated.<sup>[9]</sup> Furthermore, a mathematical

model can be built and validated so that new pathways or components can be rapidly engineered and tested *in silico*.

The carboxylic acid reductases (CARs) are enzymes that have increasing interest as biocatalysts.<sup>[10,11]</sup> CARs catalyze the reduction of carboxylic acids to aldehydes in mild conditions, and can connect many types of enzyme reaction allowing the construction of novel multi-enzyme pathways (Figure 1A).<sup>[12–14]</sup> Previously, we have shown CARs to be a fairly promiscuous enzyme class, catalyzing the reduction of both aliphatic and aromatic carboxylic acids. Electron rich acids are favored as the first step in the CAR reaction mechanism, attack by the carboxylate on the  $\alpha$ -phosphate of ATP, is limiting.<sup>[15]</sup> CARs have been shown to be useful in the construction of pathways for *in vivo* use. Some examples include the production of the flavor vanillin by yeast<sup>[16]</sup> and a synthetic pathway for the production of propane in *Escherichia coli*.<sup>[17]</sup> The use of CARs as an isolated enzyme in a multistep reaction, however, has not been explored. Possibly this is due to the challenging and costly requirements for both ATP and NADPH regeneration,<sup>[18]</sup> or the buildup of product inhibition by pyrophosphate (PP<sub>i</sub>).<sup>[15,19]</sup>

Many oxidoreductases require cofactors, most commonly the comparatively expensive NAD(P)H. For biocatalysis to be an economically viable process *in vitro*, cofactor regeneration is essential. Many enzyme systems have been developed for the regeneration of NAD(P)H.<sup>[20]</sup> One such system is the phosphite dehydrogenase (PTDH) enzyme from *Pseudomonas stutzeri*. The wild-type enzyme regenerates NADH via the nearly irreversible oxidation of phosphite to phosphate. The enzyme has been engineered to regenerate NADPH and improve its thermostability,<sup>[21,22]</sup> making it more attractive when regeneration of both NADH and NADPH is required.

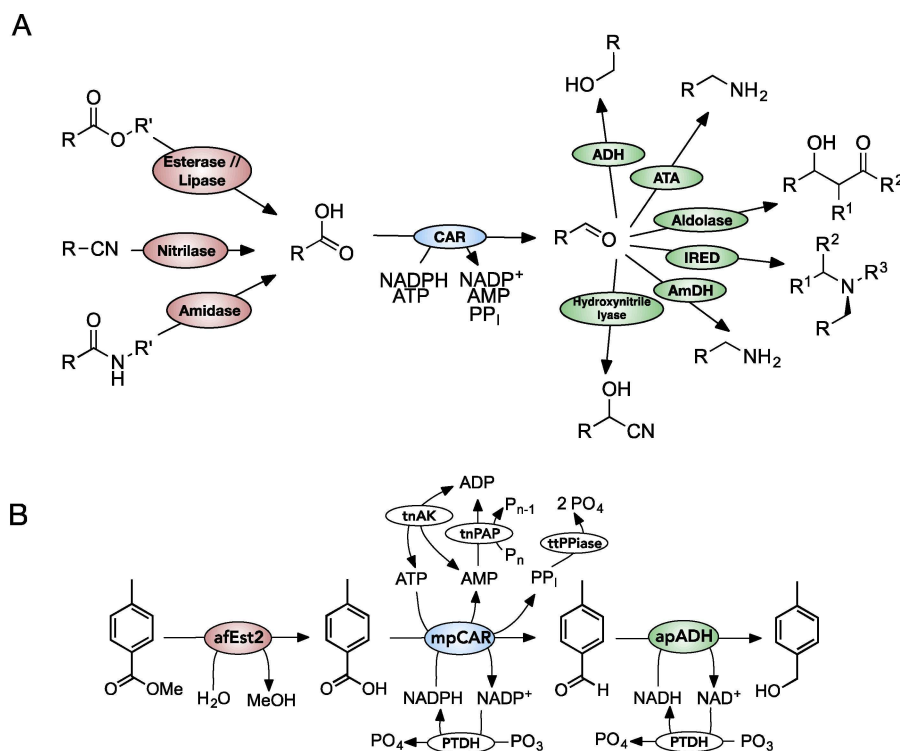
Techniques for the regeneration of ATP are less well developed. Enzymes such as pyruvate kinase (PK), creatine kinase (CK), adenylate kinase (AK) and polyphosphate kinase (PPK) have been used.<sup>[23]</sup> Of these, polyphosphate kinase makes use of the cheapest and most stable substrate, polyphosphate.

[a] Dr. W. Finnigan, Prof. J. A. Littlechild, Prof. N. J. Harmer  
Department of Biosciences  
University of Exeter  
Henry Wellcome Building for Biocatalysis  
Stocker Road, Exeter EX4 4QD (UK)  
E-mail: N.J.Harmer@exeter.ac.uk

[b] R. Cutlan, Prof. N. J. Harmer  
Living Systems Institute  
University of Exeter  
Stocker Road, Exeter EX4 4QDm (UK)

[c] Dr. R. Snajdrova, Dr. J. P. Adams  
GlaxoSmithKline R&D Ltd  
Medicines Research Centre  
Gunnels Wood Road, Stevenage  
Hertfordshire SG1 2NY (UK)

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cctc.201900646>



**Figure 1.** Utilizing CARs in multi-step enzyme reactions. A: CARs join many industrially relevant enzyme reactions, making them useful for the construction of novel multi-step enzyme reactions. Cofactors and additional substrates are not shown for enzymes other than CAR for clarity. ADH: alcohol dehydrogenase, ATA: amino transferase, IRED: imine reductase, AmDH: amine dehydrogenase, CAR: carboxylic acid reductase. R: R-group, limited by enzyme substrate specificity. B: A schematic of the seven enzyme reaction. The hydrolysis of methyl 4-toluate to 4-toluic acid, followed by reduction to 4-tolualdehyde and further to 4-tolylalcohol is shown. The use or production of water is not shown. afEst2: Esterase enzyme from *Archaeoglobus fulgidus*, mpCAR: Carboxylic acid reductase from *Mycobacterium phlei*, apADH: Alcohol dehydrogenase from *Aeropyrum pernix*, PTDH: Engineered phosphite dehydrogenase from *Pseudomonas stutzeri*, ttPPIase: Inorganic pyrophosphatase from *Thermus thermophilus*, tnPAP: Polyphosphate AMP phosphotransferase from *Thermodesulfobium narugense*, tnAK: Adenylate kinase from *Thermotoga neapolitana*. PolyP<sub>n</sub>: A polyphosphate molecule with a chain length of n phosphates.

Polyphosphate is also a substrate for polyphosphate-AMP phosphotransferase (PAP), which allows the regeneration of ADP from AMP.<sup>[24]</sup> The combination of PAP and PPK therefore allowed complete regeneration of ATP from AMP.<sup>[25,26]</sup> In an alternative approach, adenylate kinase (which catalyzes the reversible phosphorylation of AMP by ATP) was used in place of PPK. Coupling of this enzyme with PAP pushes the equilibrium towards ATP regeneration (Figure 1B).<sup>[27,28]</sup>

Synthetic biology has long promised the use of well-characterized parts for the rational design of new metabolic pathways.<sup>[29–31]</sup> However the use of modelling to optimize these *in vivo* reactions is yet to be fully realized.<sup>[32]</sup> In contrast, modelling of enzymes *in vitro* can be fairly robust and offers solutions for reaction engineering.<sup>[33]</sup> This could allow the combination of enzymes, for which validated mathematical models exist, to be rapidly engineered and tested *in silico*.<sup>[2]</sup>

Indeed, in developing new biocatalytic processes the use of kinetic modelling is widely advocated, yet often not used in process development.<sup>[34,35]</sup> The development of a kinetic model early on in the development process can be invaluable in cost/benefit or feasibility analysis.<sup>[35]</sup> It permits evidence-based decision-making,<sup>[36]</sup> and critically allows for the identification of bottlenecks and the quantification of process problems (e.g. feedback inhibition or inhibition by side-products).<sup>[6]</sup> Mechanis-

tic models, which seek to describe enzyme mechanisms as accurately as possible, seek to understand a system as well as to predict it. These offer opportunities to develop substantial improvements or insights into the development process.<sup>[33]</sup> A classic example is the use of Michaelis-Menten equations.<sup>[37]</sup> Such integrated kinetic models of multi-enzyme processes are challenging because of the large number of kinetic parameters involved.<sup>[36]</sup> Deterministic models for enzyme processes with two steps (plus cofactor regeneration) have been explored,<sup>[36,38]</sup> alternative models have fitted kinetic parameters based on product concentrations.<sup>[39,40]</sup> The most ambitious of these fitted parameters is for up to fifteen enzymes.<sup>[41]</sup>

To demonstrate the use of CARs in multi-step cascade reactions, and investigate their use *in vitro*, we designed a reaction made up of an esterase, a CAR and an alcohol dehydrogenase (ADH) to hydrolyze and then reduce an ester to its corresponding alcohol. Methyl 4-toluate was chosen as a trial substrate. Whilst not directly industrially relevant, it acts as a good model for many industrially relevant compounds. The three step reaction models chemically simple reactions to supply an acid to the CAR, and to utilize the aldehyde product. Enzymes for cofactor regeneration and removal of an inhibitory by-product were added to provide an efficient reaction with sub-stoichiometric cofactor concentrations. We demonstrate

that this results in an effective three step cascade for the ester to alcohol transformation incorporating seven enzymes over eight biochemical steps.

To highlight the modular nature of the enzyme toolbox for designing multistep reactions, we characterized each enzyme individually and established a mathematical model for the entire pathway. In constructing and testing the model for each step in the reaction we demonstrate the value of building a model in identifying process problems. Specifically, we were able to predict the need for PP<sub>i</sub> removal, and a non-enzymatic reaction of an intermediate with a component of the reaction. We then performed optimization of the trial batch reaction, and achieved the target 90% yield with the minimal concentration of the enzymes. Our results demonstrate that a deterministic model can be robust for enzyme cascades with cofactor recycling involving at least seven enzymes. We further demonstrate that these models have great potential to quickly assemble novel enzyme catalyst networks.<sup>[42]</sup>

## Results

### Expression and Purification of Enzymes

All of the enzymes used in this study were recombinantly prepared from *E. coli*, as fusions with a polyhistidine-tag. All of the proteins were purified by nickel affinity chromatography followed by size exclusion chromatography. Each enzyme was purified to >90% purity (Supplementary Figure 1 to Supplementary Figure 8).

### Designing a Synthetic Multi-Step Pathway

Enzymes with the correct substrate specificity to catalyze the esterase, CAR and ADH steps were identified from the literature. Where possible, thermostable enzymes were chosen to provide maximum operational stability. In the case of the CAR step, only moderate thermostability was possible due to the limited number of organisms that use these enzymes. An esterase from the hyperthermophile *Archaeoglobus fulgidus* (afEst2), a CAR from the moderate thermophile *Mycobacterium phlei* (mpCAR), and an ADH from the hyperthermophile *Aeropyrum pernix* (apADH) were chosen (Figure 1B).<sup>[43,44]</sup> The catalytic constants of these enzymes against the relevant substrates in the test pathway were determined (Supplementary Figure 19 to Supplementary Figure 38) or taken from relevant literature (references shown in Table 1).

To allow cofactor regeneration, a thermostable mutant of PTDH, capable of regenerating both NADH and NADPH, was chosen given the need for both of these cofactors in the pathway. For ADP regeneration from AMP, we identified a PAP enzyme from the thermophile *Thermodesulfobium narugense* (tnPAP) with 33% identity to the previously characterized PAP from *Acinetobacter johnsonii*.<sup>[45]</sup>

A thermostable PPT enzyme from *Thermosynechococcus elongatus* that had previously been characterized<sup>[46]</sup> was initially

Table 1. Kinetic parameters.<sup>[a]</sup>

afEst2 <sup>[44]</sup>	
$k_{cat-Fwd}$ <sup>[44]**</sup>	9.3 ± 4.65 or 12 ± 6 min <sup>-1</sup>
$K_{M-Ester}$ <sup>[44]</sup>	1,500 ± 375 μM
$K_{M-H2O}$ <sup>**</sup>	1,000–100,000 μM
$K_{cat-Rev}$ <sup>**</sup>	1–50 min <sup>-1</sup>
$K_{M-MeOH}$ <sup>**</sup>	1,000–100,000 μM
$K_{M-Acid}$ <sup>**</sup>	100–100,000 μM
mpCAR <sup>[15]</sup>	
$k_{cat}$	200 ± 20 min <sup>-1</sup>
$K_{M-Acid}$	1,500 ± 320 μM
$K_{M-ATP}$	100 ± 28 μM
$K_{I-ATP}$	40 ± 34 μM
$K_{M-NADPH}$	30 ± 8 μM
$K_{I-AMP}$ <sup>[15]</sup>	10,000 ± 1,800 μM
$K_{I-ADP}$ <sup>[15]</sup>	11,000 ± 4,000 μM
$K_{I-NADP+}$ <sup>[15]</sup>	143 ± 16 μM
$K_{I-PPi-Acid}$ <sup>[15]</sup>	340 ± 80 μM
$K_{I-PPi-ATP}$ <sup>[15]</sup>	220 ± 100 μM
$\alpha_{-PPi-ATP}$ <sup>[15]</sup>	2.6 ± 2.8
apADH	
$k_{cat-Fwd}$	1.7 ± 0.2 min <sup>-1</sup>
$K_{cat-Rev}$ <sup>*</sup>	1.7 ± 0.85 min <sup>-1</sup>
$K_{M-NADH}$	180 ± 60 μM
$K_{M-NAD+}$	190 ± 40 μM
$K_{I-NAD+}$ <sup>*</sup>	185 ± 92.5 μM
$K_{I-NADH}$ <sup>*</sup>	185 ± 92.5 μM
$K_{M-Aldehyde}$	350 ± 120 μM
$K_{M-Alcohol}$	10,000 ± 5,000 μM
ttPpiase <sup>[55]</sup>	
$k_{cat}$	4,400 ± 2,200 min <sup>-1</sup>
$K_{M-PPi}$ <sup>[55]</sup>	500 ± 250 μM
PTDH	
$K_{cat-NAD+}$	637 ± 16 min <sup>-1</sup>
$K_{cat-NADP+}$	342 ± 16 min <sup>-1</sup>
$K_{M-NAD+}$	85 ± 5 μM
$K_{M-NADP+}$	220 ± 40 μM
tnPAP	
$k_{cat-Fwd}$	250 ± 125 min <sup>-1</sup>
$K_{M-AMP}$	280 ± 240 μM
$K_{M-PolyP}$	4,000 ± 2,000 μM
$K_{cat-Rev}$ <sup>[56]</sup>	3.4 ± 1.7 min <sup>-1</sup>
$K_{M-ADP}$ <sup>[56]</sup>	8,300 ± 4,150 μM
tnAK <sup>[47,57]</sup>	
$K_{cat-ADP}$ <sup>[47]</sup>	2,340 ± 1,170 min <sup>-1</sup>
$K_{cat-AMP-ATP}$ <sup>[47]</sup>	3,950 ± 1,975 min <sup>-1</sup>
$K_{M-ADP}$ <sup>[57]</sup>	91 ± 45.5 μM
$K_{M-AMP}$ <sup>[57]</sup>	38 ± 19 μM
$K_{M-ATP}$ <sup>[57]</sup>	51 ± 25.5 μM
Aldehyde side reaction	
$k$	0.00279 ± 0.001395 min <sup>-1</sup>
[a] Parameters were determined experimentally or obtained from the literature. Errors represent 95% confidence intervals where these could be experimentally determined or obtained from the literature. Where these could not be experimentally determined, errors of 50% of the parameter value were conservatively assigned. In (a few) cases a reasonable estimate had to be made, marked with *. Where parameters were completely unknown, large uncertainty bounds were used, marked with **. *** For the esterase alone, and in the complete reaction respectively, reflecting data in Supplementary Figure 16.	

**Table 2.** Kinetic equations. The rationale for the kinetic equations chosen for each enzyme and interacting partner are discussed in detail below. Aldehyde degradation was modelled as a first order process.

afEst2	
$r1 = C_{Est} \cdot k_{cat}^{Est} \cdot \frac{C_{Ester} \cdot C_{H2O}}{(K_M^{H2O} \cdot C_{Ester}) + (K_{Ester}^{Ester} \cdot C_{H2O}) + (C_{Ester} \cdot C_{H2O})}$	(1)
$r2 = C_{Est} \cdot k_{cat}^{Est} \cdot \frac{C_{Acid} \cdot C_{Methanol}}{(K_M^{MethOH} \cdot C_{Acid}) + (K_{Acid}^{Acid} \cdot C_{Methanol}) + (C_{Acid} \cdot C_{Methanol})}$	(2)
mpCAR	
$K_M^{ATP} = K_M^{ATP} \cdot \frac{1 + \frac{C_{AMP}}{K_{ATP}^{AMP}}}{1 + \frac{C_{ATP}}{K_{ATP}^{ATP}}} \cdot (1 + \frac{C_{AMP}}{K_{ATP}^{AMP}}) \cdot (1 + \frac{C_{ADP}}{K_{ATP}^{ADP}})$	(3)
$K_M^{NADPH} = K_M^{NADPH} \cdot (1 + \frac{C_{NADP+}}{K_{NADPH}^{NADP+}})$	(4)
$K_M^{Acid} = K_M^{Acid} \cdot (1 + \frac{C_{PPi}}{K_{PPi}^{PPi}})$	(5)
$K_{cat}^{CAR} = \frac{K_{cat}^{CAR}}{1 + \frac{C_{PPi}}{K_{PPi}^{PPi}}}$	(6)
$r3 = C_{CAR} \cdot K_{cat}^{CAR} \cdot \frac{C_{Acid} \cdot C_{ATP} \cdot C_{NADPH}}{(K_{ATP}^{ATP} \cdot K_M^{Acid}) + (K_M^{NADPH} \cdot C_{ATP} \cdot C_{Acid}) + (K_M^{Acid} \cdot C_{ATP} \cdot C_{NADPH}) + (K_{ATP}^{ATP} \cdot C_{Acid} \cdot C_{NADPH}) + (C_{ATP} \cdot C_{Acid} \cdot C_{NADPH})}$	(7)
apADH	
$r4 = C_{ADH} \cdot k_{cat}^{ADHFwd} \cdot \frac{C_{Aldehyde} \cdot C_{NADH}}{(K_M^{NADH} \cdot K_M^{Aldehyde}) + (K_M^{Aldehyde} \cdot C_{NADH}) + (C_{NADH} \cdot C_{Aldehyde}) + (C_{NADH} \cdot C_{Aldehyde} \cdot C_{NAD+})}$	(8)
$r5 = C_{ADH} \cdot k_{cat}^{ADHRev} \cdot \frac{C_{Alcohol} \cdot C_{NAD+}}{(K_{NAD+}^{NAD+} \cdot K_M^{Alcohol}) + (K_M^{Alcohol} \cdot C_{NAD+}) + (K_M^{NAD+} \cdot C_{Alcohol}) + (C_{NAD+} \cdot C_{Alcohol})}$	(9)
PTDH	
$K_M^{NAD+} = K_M^{NAD+} \cdot (1 + \frac{C_{NADP+}}{K_{NADP+}^{NADP+}})$	(10)
$K_M^{NADP+} = K_M^{NADP+} \cdot (1 + \frac{C_{NAD+}}{K_{NAD+}^{NAD+}})$	(11)
$r6 = C_{PTDH} \cdot k_{cat}^{PTDH-NAD+} \cdot \frac{C_{NAD+}}{C_{NAD+} + K_M^{NAD+}}$	(12)
$r7 = C_{PTDH} \cdot k_{cat}^{PTDH-NADP+} \cdot \frac{C_{NADP+}}{C_{NADP+} + K_M^{NADP+}}$	(13)
PPiase	
$r8 = C_{PPiase} \cdot k_{cat}^{PPiase} \cdot \frac{C_{PPi}}{C_{PPi} + K_M^{PPi}}$	(14)
PAP	
$r9 = C_{PPT} \cdot k_{cat}^{PPTFwd} \cdot \frac{C_{PolyP} \cdot C_{AMP}}{C_{PolyP} + K_M^{PolyP} \cdot (C_{AMP} + K_M^{AMP})}$	(15)
$r10 = C_{PPT} \cdot k_{cat}^{PPTRev} \cdot \frac{C_{ADP}}{C_{ADP} + K_M^{ADP}}$	(16)
AK	
$r11 = C_{AK} \cdot k_{cat}^{AK-Fwd} \cdot \frac{C_{ADP}}{C_{ADP} + K_M^{ADP}} \cdot \frac{C_{ADP}}{C_{ADP} + K_M^{ADP}}$	(17)
$r12 = C_{AK} \cdot k_{cat}^{AK-Rev} \cdot \frac{C_{AMP}}{C_{AMP} + K_M^{AMP}} \cdot \frac{C_{ATP}}{C_{ATP} + K_M^{ATP}}$	(18)
Aldehyde degradation	
$r13 = k^{ad} \cdot C^{Aldehyde}$	(19)

chosen to regenerate ATP from ADP. However, in our hands this enzyme gave very low activity. In its place, a thermostable AK enzyme from *Thermotoga neapolitana* (tnAK) was used, which had previously been characterized with high activity.<sup>[47]</sup> We had previously identified that PP<sub>i</sub> is a significant inhibitor of CARs.<sup>[15]</sup> We therefore also determined the activity of a thermostable PPiase from *Thermus thermophilus* (ttPPiase; Supplementary Figure 33).

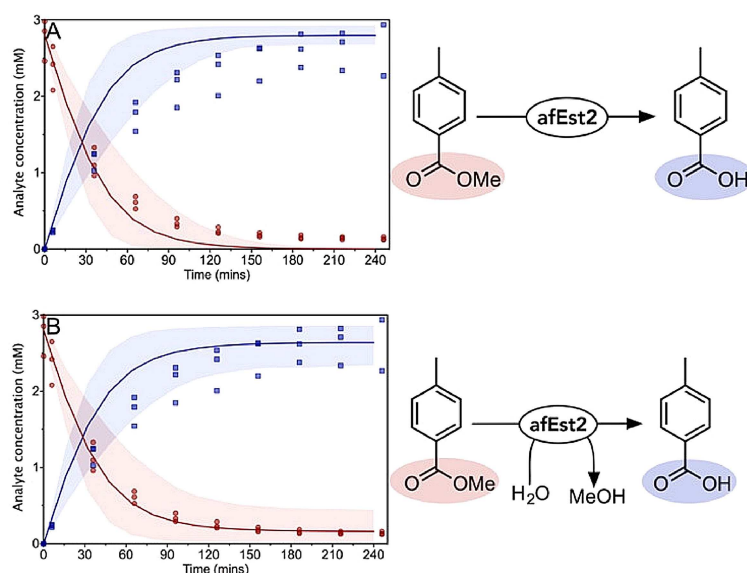
### Identifying an Operational Window

Thermostability and activity at different pHs were determined for each enzyme to define an operational window for the reaction (Supplementary Figure 39). Data for afEst2,<sup>[44]</sup> mpCAR,<sup>[15]</sup> apADH<sup>[48]</sup> and tnAK<sup>[47]</sup> were adapted from previous work. The effects of pH and temperature on the remaining enzyme activities were characterized (Supplementary Figures 11

to 14). A pH of 7.5 was chosen as a compromise for all enzymes, and ADH activity in the (undesired) oxidative direction was minimized at this pH. As we expected, the operational window for temperature is primarily limited by the mpCAR enzyme, as this has only moderate thermostability. A reaction temperature of 30 °C was chosen to ensure maximum activity of this enzyme.

### Reaction Modelling

Mathematical models utilizing kinetics based on the steady state approximation were developed in isolation for each section of the multi-step reaction and validated before being combined into a multi-step process. These models were developed using Python scripts with equations appropriate for each enzyme. The rate equations are summarized in Table 2, which were used to construct differential equations for each reaction to be tested (Supplementary Table 2). Each rate equation is based upon the



**Figure 2.** Validation of the esterase model. A reaction containing 10  $\mu\text{M}$  afEst2 and 2,800  $\mu\text{M}$  methyl 4-toluate (ester) was used to validate the model for afEst2. Ester and acid concentrations, measured every 30 minutes by HPLC, are shown as red circles and blue squares respectively. The model prediction is shown as the solid line in the same colors.  $\chi^2 = 2.9, 6.3, 32$  for the deterministic, quadratic and linear models respectively; when the reverse reaction is included in the deterministic model,  $\chi^2$  reduces to 0.82. The shaded areas represent the values between the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Data show three experimental replicates for each point. A: Esterase reaction modelled as an irreversible, one substrate reaction. B: Esterase reaction modelled as a reversible bi-bi reaction, with estimated parameters for the reverse direction.

steady-state approximation, and represents a unidirectional reaction.<sup>[49,50]</sup> Where reactions are reversible, equations for both directions were included. The rationale for each equation used is detailed below. Kinetic parameters were either identified from the available literature or determined experimentally. An uncertainty analysis was carried out for each model, where possible, using bounds equal to the 95% confidence intervals of each parameter (Table 1). In some cases where uncertainty was judged to be large due to incomplete knowledge, bounds of  $\pm 50\%$  of the parameter were used. Starting concentrations were given bounds of  $\pm 5\%$ , except in the case of polyphosphate where the length of the polyphosphate chain is unknown. In this case  $\pm 25\%$  was used with the upper bound equal to the absolute concentration of phosphate units. Model predictions were tested by running small scale reactions in a thermomixer, with samples taken every 30 minutes and quenched with acetonitrile for analysis by HPLC.

### Esterase Reaction

The afEst2 has been recently characterized and kinetic parameters for the hydrolysis of methyl 4-toluate at 30 °C reported<sup>[44]</sup> (Table 1). As this characterization was carried out at pH 8.2, a medium level of  $\pm 25\%$  uncertainty was associated with these parameters. Hydrolysis reactions are typically modelled using the irreversible, one substrate Michaelis-Menten equation. This model was tested against a small scale batch reaction (Figure 2A), which it was able to make a reasonable prediction. A reduced  $\chi^2$  statistic was used to assess the model.<sup>[51]</sup> A model fitted directly to the experimental data would be expected to achieve a  $\chi^2$  of one; we interpreted models as strong where  $\chi^2$  was less than fifteen, and less than a

quadratic or simple linear model fitted to the experimental data. Here,  $\chi^2 = 2.9$  (c.f. 6.3 and 32 for the alternative models), and so was interpreted as a good fit.

However, on examination of the data, it is noticeable that the reaction proceeds faster than modelled, and that the reaction did not appear to go to completion. The characterization of this enzyme was performed in 2.5 mM buffer (necessarily as the assay detects a change in pH),<sup>[44]</sup> whilst all of our reactions used 100 mM buffer. We tested the effect of changing the buffer concentration on afEst2, and found that the enzyme is 55% faster in 100 mM buffer (Supplementary Figure 15). To reflect this the  $k_{\text{cat-Fwd}}$  was increased by 55% to 9.3  $\text{min}^{-1}$ , with  $\pm 50\%$  uncertainty. While an irreversible reaction was initially assumed due to an excess of water, we calculated the  $K_{\text{eq}}$  for this reaction as 22 (Supplementary Table 1), consistent with a noticeable reverse reaction. We therefore incorporated the reverse direction into our model (Table 2), with estimated parameters with large uncertainty bounds (Table 1).<sup>[52-54]</sup> The updated model predicted the batch reaction well, albeit with a larger degree of uncertainty (Figure 2B). A sensitivity analysis was carried out to look at the sources of uncertainty in this new model (Supplementary Figure 16). The revised model showed  $\chi^2 = 0.81$ , highlighting the improvement of the model.

### CAR Reaction

The CAR enzymes have three substrates: ATP, NADPH and a carboxylic acid. The reaction proceeds in an ordered fashion, with the ATP activating the acid and an enzyme-acid conjugate being formed before NADPH binds.<sup>[58]</sup> This reaction was there-

fore modelled using a three substrate rate equation, analogous to that of an aminoacyl tRNA synthetase in which a ternary complex must form before the third substrate can bind.<sup>[59]</sup> Parameters were determined experimentally (Supplementary Figures 21 and 22).

A reaction to test the CAR model was set up, predicting a fast turnover of the acid into its derivative aldehyde (shown in grey, Figure 3A). Instead, after an initial burst, activity slowed considerably and the reaction was not complete even after four hours ( $\chi^2=200$ ). We investigated the possibility of product inhibition and found that all of the products of the CAR reaction ( $PP_i$ , AMP and  $NADP^+$ ) act as inhibitors (discussed in more detail in [15]). While AMP and  $NADP^+$  both act as competitive inhibitors which may be overcome by increasing the substrate concentration,  $PP_i$  acts as a mixed model inhibitor with respect to ATP, causing a large decrease in the apparent  $k_{cat}$ . When the inhibition by these products was taken into account the model fitted the data quite well (Figure 3A;  $\chi^2=3.1$ ). We noted that modelling the inhibition resulted in a large level of uncertainty.

### ttPPIase Removes CAR Inhibition by $PP_i$

In order to alleviate inhibition by  $PP_i$  on the CAR enzyme, an inorganic pyrophosphatase from *T. thermophilus* (ttPPIase) was added to the reaction. This enzyme shows exceptional thermostability and retained over 90% activity after a 30 minute incubation at 95 °C (Supplementary Figure 40). Pyrophosphatase activity was measured at a saturating concentration of  $PP_i$  (5 mM) and the rate taken to be the  $k_{cat}$ . The enzyme was modelled using the one substrate Michaelis-Menten equation, and using the  $K_M$  determined in a recent study on this enzyme.<sup>[60]</sup> The addition of the ttPPIase alleviated the majority of the inhibitory effects seen previously and the CAR reaction went to completion within one hour (Figure 3B).

### Aldehyde Side Reaction

Once the CAR reaction was complete, aldehyde concentration decreased over time (Figure 3B). Aldehydes can react with free amines; and benzaldehyde in particular has a documented specific reaction with Tris (Supplementary Figure 17).<sup>[61,62]</sup> We verified that 4-methylbenzaldehyde reacts in a similar manner with Tris, and validated the product by mass spectrometry (Supplementary Figure 18). As we validated this side reaction after completing all other experiments, we accounted for this reaction using a one phase decay equation. This was fitted to the decrease in aldehyde concentration, and an equation for the rate of aldehyde side reaction constructed (Table 1 and Table 2). This resulted in a very good fit of the aldehyde concentration to the data ( $\chi^2=2.4$ ).

### NAD(P)H Regeneration Using PTDH

The regeneration of NADPH using PTDH was then added to the CAR step. Although PTDH shows sequential ordered kinetics with  $NAD^+$  binding before phosphite, the effect of phosphite concentration was not modelled as all experiments were performed at saturating phosphite concentrations. The one substrate Michaelis-Menten equation was therefore used for reactions with either

$NAD^+$  or  $NADP^+$  as the substrate.  $NAD^+$  and  $NADP^+$  act as competitive inhibitors of each other with  $K_i$  values equal to their respective  $K_M$  values, and this inhibition was added to the model (Table 2). Kinetic parameters for  $NAD^+$ ,  $NADP^+$  were determined at assay conditions by following the generation of NADH or NADPH at 340 nm (Table 1; Supplementary Figures 31 and Figure 32).

The NADPH regeneration was tested by attempting reduction of 4,000  $\mu$ M 4-toluate by CAR with only 500  $\mu$ M or 50  $\mu$ M NADPH, together with the PTDH regeneration system. Both reactions went to completion with far less than stoichiometric NADPH. The reaction containing 500  $\mu$ M NADPH fitted the model predictions reasonably well (Figure 3C;  $\chi^2=160$ , in this case driven by a single anomalous data point). When only 50  $\mu$ M NADPH was used, a much slower reaction was predicted by the model. However, only a small difference was observed experimentally compared to the use of 500  $\mu$ M NADPH (Figure 3D;  $\chi^2=1,200$ ). We considered that some form of substrate channeling<sup>[63,64]</sup> may be taking place between the CAR and PTDH resulting in a lower apparent  $K_M$  for  $NADP^+$  by PTDH.

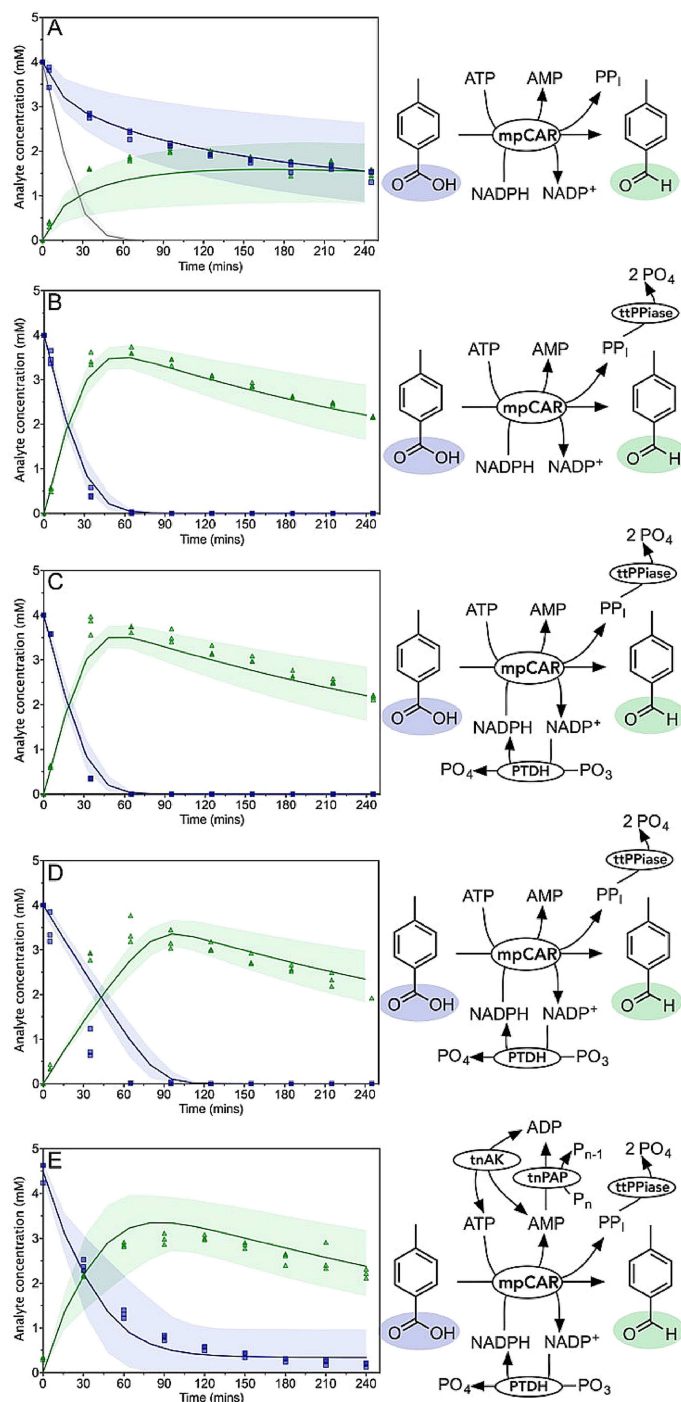
### ATP Regeneration Using tnPAP and tnAK

tnPAP showed good PAP activity, and kinetic parameters for polyphosphate and AMP were determined experimentally (Supplementary Figures 35 to 37), and a bi-substrate equation constructed. An adenylate kinase from *Thermotoga neapolitana* has been reported to possess excellent thermostability, with kinetic parameters similar to those of the characterized *E. coli* adenylate kinase. These parameters were used to model the adenylate kinase. However, testing of this enzyme in a CAR coupled reaction revealed a significantly lower turnover number (Supplementary Figure 39). Furthermore, the magnesium concentration has been shown to be instrumental in controlling the equilibrium catalyzed by AK. Since the free magnesium concentration is not known due to chelation by polyphosphate and other compounds, this effect was not taken into account. For this reason, the AK reaction is difficult to model accurately in the context of the multistep pathway. Therefore, a simple bi-substrate equation was used to describe it. Large uncertainty bounds were used for all the parameters in the AK reaction as a consequence. With all the cofactor regeneration systems in place, the reaction proceeded almost to completion with far less than stoichiometric cofactors, and fitted well to the model (Figure 3E;  $\chi^2=11$ ).

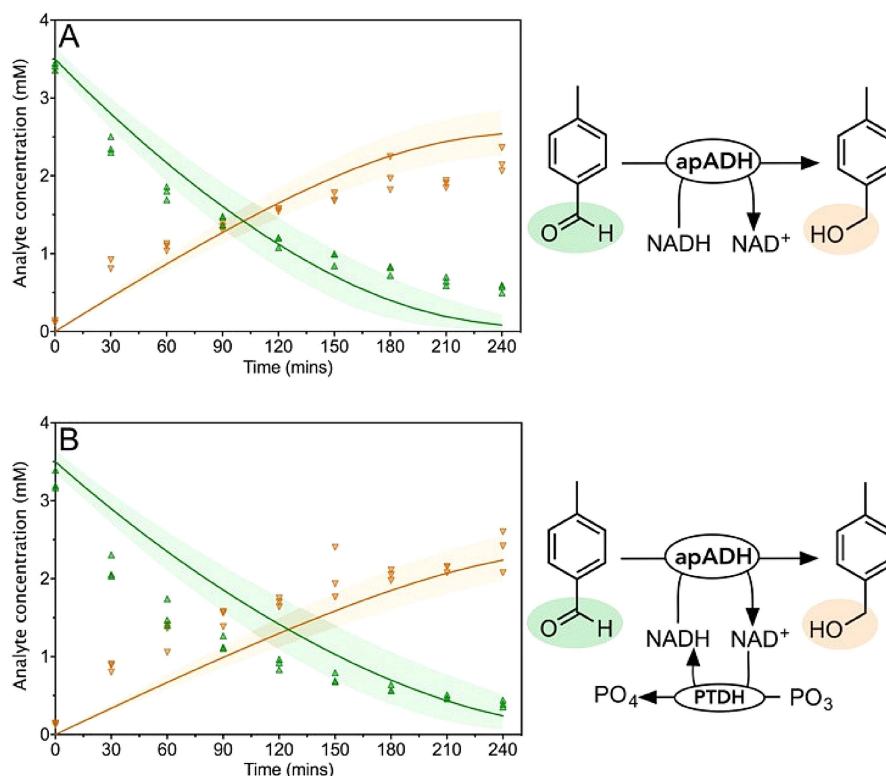
### ADH Reaction

The kinetic parameters for both the forward and reverse directions of the apADH catalyzed reactions were determined experimentally. The  $K_M$  for 4-tolyl alcohol (for the reverse oxidative reaction) could not be accurately determined, as the reaction is not saturated at the highest concentrations that the substrate is soluble at (Supplementary Figures 29 and 30). An estimated  $K_M$  of 100 mM was used (Table 1). Consequently,  $k_{cat-Rev}$  was also difficult to determine and a value similar to  $k_{cat-Fwd}$  was assumed. A two substrate steady state sequential rate equation was used to describe both the forward and reverse reactions separately, using an estimated parameter for the  $K_i$  for  $NAD^+$  and NADH based on their determined  $K_M$  values.

Reactions featuring only apADH as well as with PTDH for NADH regeneration (using far less than stoichiometric NADH) were used to validate the model, which predicted the rate of reduction well in both cases (Figure 4;  $\chi^2=7.0$ ). A sensitivity analysis of the final alcohol concentration in both these reactions showed the  $K_i$  for



**Figure 3.** Validation of mpCAR and related enzyme models. Reactions were performed to validate the model for mpCAR, and the addition of cofactor regenerating and PP<sub>i</sub> removing enzymes. Acid and aldehyde concentrations measured every 30 minutes by HPLC are shown as blue squares and green upwards facing triangles respectively. The model prediction is shown as the solid line in the same colors. The shaded areas represent the values between the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Data show three experimental replicates for each point. A: mpCAR alone. Reactions were initiated with 1 μM mpCAR, 8,000 μM ATP, 5,000 μM NADPH, 20,000 μM MgCl<sub>2</sub>, 4,000 μM 4-toluic acid. The initial model prediction, which did not take into account any product inhibition, is shown in grey. B: mpCAR-ttPpiase. Reactions were initiated with 1 μM mpCAR, 1 μM ttPpiase, 8,000 μM ATP, 5,000 μM NADPH, 20,000 μM MgCl<sub>2</sub>, 4,000 μM 4-toluic acid. C: mpCAR-ttPpiase-PTDH. Reactions were initiated with 1 μM mpCAR, 1 μM ttPpiase, 1 μM PTDH, 8,000 μM ATP, 500 μM NADPH, 20,000 μM MgCl<sub>2</sub>, 20,000 μM PO<sub>3</sub>, 4,000 μM 4-toluic acid. D: mpCAR-ttPpiase-PTDH (low NADPH). Reactions were initiated with 1 μM mpCAR, 1 μM ttPpiase, 1 μM PTDH, 8,000 μM ATP, 50 μM NADPH, 20,000 μM MgCl<sub>2</sub>, 20,000 μM PO<sub>3</sub>, 4,000 μM 4-toluic acid. E: mpCAR-ttPpiase-PTDH-tnPAP-tnAK. Reactions were initiated with 0.4 μM mpCAR, 1 μM ttPpiase, 1 μM PTDH, 3 μM tnPAP, 1 μM tnAK, 1,250 μM ATP, 500 μM NADPH, 20,000 μM MgCl<sub>2</sub>, 20,000 μM PO<sub>3</sub>, 6,000 μM polyphosphate, 4,000 μM 4-toluic acid. Observed reduced  $\chi^2$  values (for deterministic, fitted quadratic and fitted linear models): A: 3.1, 2.3, 21; B: 2.4, 220, 780; C: 160, 1300, 1400; D: 1200, 10, 1100; E: 11, 30, 210.



**Figure 4.** Validation of ADH and ADH-PTDH models. Reactions used to validate the model for apADH, and cofactor regeneration by PTDH. Aldehyde and alcohol concentrations measured every 30 minutes are shown as green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colors. The shaded areas represent the values between the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Data show three experimental replicates for each point. A: apADH only. Reactions were initiated with 10  $\mu\text{M}$  apADH, 5,000  $\mu\text{M}$  NADH, 3,500  $\mu\text{M}$  4-tolualdehyde. B: apADH-PTDH. Reactions were initiated with 10  $\mu\text{M}$  apADH, 1  $\mu\text{M}$  PTDH, 20,000  $\mu\text{M}$   $\text{PO}_3$ , 500  $\mu\text{M}$  NADH, 3,500  $\mu\text{M}$  4-tolualdehyde. Observed reduced  $\chi^2$  values (for deterministic, fitted quadratic and fitted linear models): A: 13, 7.9, 45; B: 7.0, 3.9, 35.

$\text{NAD}^+$  and NADH, and  $k_{\text{cat-Rev}}$  to have little effect (Supplementary Figure 41).

### Building a Multi-Step Enzyme Cascade

Once the models for each of the esterase, CAR and ADH steps were independently validated with cofactor regeneration, multi-step reactions were constructed. The CAR step was tested in combination with first the esterase, and then ADH (Figure 5A, B;  $\chi^2 = 10, 4.8$  respectively).

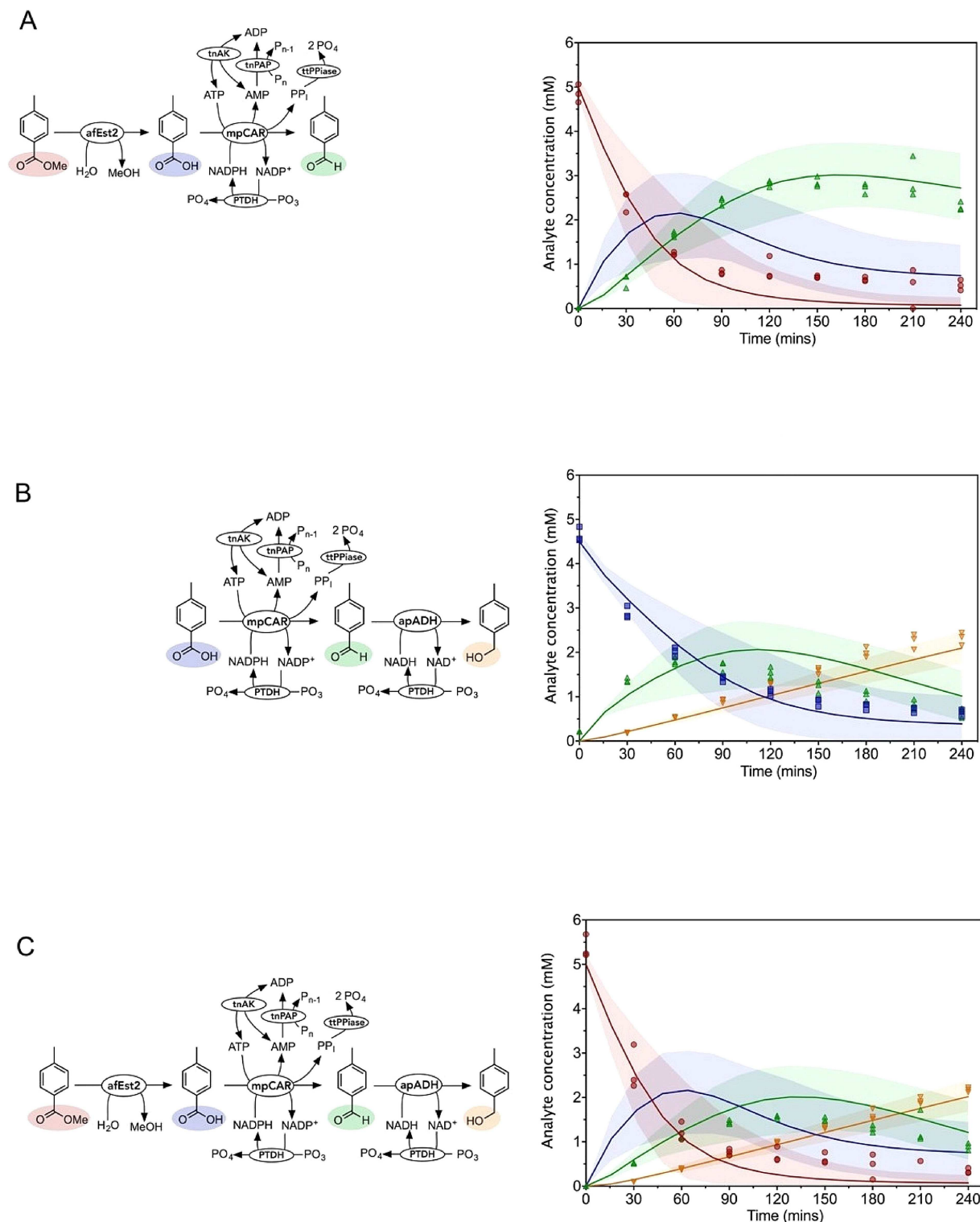
In each case the model performed well in predicting the productivity of the reaction. However, the hydrolysis reaction catalyzed by afEst2 proceeded faster than was expected (Figure 5A and 5C). We considered that one or more reagents for other reactions might increase the afEst2 activity. Addition of all of the final reaction reactants caused a 30% increase in rate, compared to the rate in 100 mM Tris (Supplementary Figure 15). Further experiments showed that no individual component accounted for this, implying a salting-in effect (Supplementary Figure 15). To accommodate this increase and the uncertainty surrounding it in the modelling, we increased the esterase  $k_{\text{cat}}$  to two times the parameter value for the combined reactions, and increased the uncertainty to 50% of this value. Higher levels of uncertainty in the afEst2 step predictions are therefore observed (Figure 5A and 5C).

### Optimization of Enzyme Concentrations by a Genetic Algorithm

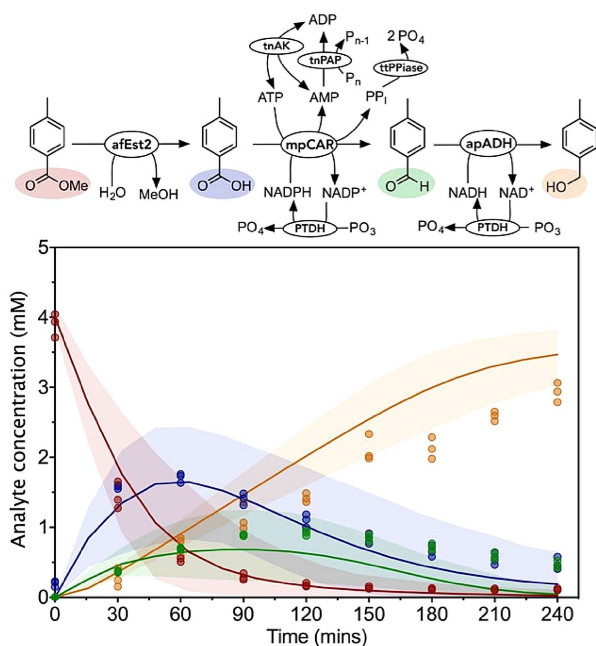
For most industrial processes, the cost of components is one critical factor in determining the economic feasibility of an approach. We therefore aimed to demonstrate that this multistep cascade could be optimized. We aimed to increase the overall yield of alcohol, whilst minimizing the overall concentration of enzymes used. We custom built a genetic algorithm (Supplementary Figure 42) to perform this optimization. Our algorithm generates random solutions for the concentration of each enzyme, and scores the quality of the solution against targets of 90% yield of alcohol, with the lowest total enzyme concentration. The highest scoring solutions are retained, and used to generate a new population of solutions. 25 cycles of this process were sufficient for the population to achieve a stable solution. Enzyme concentrations of 9.65  $\mu\text{M}$  afEst2, 0.73  $\mu\text{M}$  CAR, 24.75  $\mu\text{M}$  ADH, 0.07  $\mu\text{M}$  PPIase, 0.43  $\mu\text{M}$  PTDH 0.15  $\mu\text{M}$  AK and 0.57  $\mu\text{M}$  PAP were identified as the lowest enzyme concentration pathway capable of achieving 90% yield in four hours. This achieved a more than two-fold increase in yield, whilst reducing the concentration of six of the seven enzymes used.

To test the optimized pathway, reactions were performed at the reagent concentrations proposed by the optimization algorithm (Figure 6). The model fitted the data well ( $\chi^2 = 9.5$ ). The final yield of alcohol was slightly lower than expected reaching a final yield of only 3,000  $\mu\text{M}$  as opposed to the 3,500  $\mu\text{M}$  predicted by the model.





**Figure 5.** Validations of combinations of the esterase, CAR and ADH reactions. Reactions combining the CAR step with the esterase, ADH or both were constructed to test the model for these multi-step reactions. Ester, acid, aldehyde and alcohol concentrations measured every 30 minutes are shown as red circles, blue squares, green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colors. The shaded areas represent the values between the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Data show three experimental replicates for each point. A: afEst2-mpCAR-ttPpiase-PTDH-tnPAP-tnAK. Reactions were initiated with 10  $\mu\text{M}$  afEst2, 0.4  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 3  $\mu\text{M}$  tnPAP, 1  $\mu\text{M}$  tnAK, 5,000  $\mu\text{M}$  methyl 4-toluatoate, 500  $\mu\text{M}$  NADPH, 1,250  $\mu\text{M}$  ATP, 20,000  $\mu\text{M}$  phosphite, 6,000  $\mu\text{M}$  polyphosphate, 20,000  $\text{MgCl}_2$ . B: mpCAR-ttPpiase-PTDH-tnPAP-tnAK-apADH. Reactions were initiated with 0.4  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 3  $\mu\text{M}$  tnPAP, 1  $\mu\text{M}$  tnAK, 10  $\mu\text{M}$  apADH, 4,500  $\mu\text{M}$  4-toluic acid, 500  $\mu\text{M}$  NADPH, 500  $\mu\text{M}$  NADH, 1,250  $\mu\text{M}$  ATP, 20,000  $\mu\text{M}$  phosphite, 6,000  $\mu\text{M}$  polyphosphate, 20,000  $\text{MgCl}_2$ . C: afEst2-mpCAR-ttPpiase-PTDH-tnPAP-tnAK-apADH. Reactions were initiated with 10  $\mu\text{M}$  afEst2, 0.4  $\mu\text{M}$  mpCAR, 1  $\mu\text{M}$  ttPpiase, 1  $\mu\text{M}$  PTDH, 3  $\mu\text{M}$  tnPAP, 1  $\mu\text{M}$  tnAK, 10  $\mu\text{M}$  apADH, 5,000  $\mu\text{M}$  methyl 4-toluatoate, 500  $\mu\text{M}$  NADPH, 500  $\mu\text{M}$  NADH, 1,250  $\mu\text{M}$  ATP, 20,000  $\mu\text{M}$  phosphite, 6,000  $\mu\text{M}$  polyphosphate, 20,000  $\text{MgCl}_2$ . Observed reduced  $\chi^2$  values (for deterministic, fitted quadratic and fitted linear models): A: 7.8, 150, 5200; B: 4.8, 8.9, 47; C: 5.4, 7.5, 24.



**Figure 6.** Validations of the optimized complete reaction. A reaction using enzyme concentrations predicted by a genetic algorithm to give the lowest cost reaction, whilst hitting a target of 90% alcohol yield, was performed. Ester, acid, aldehyde and alcohol concentrations measured every 30 minutes are shown as red circles, blue squares, green upwards facing triangles and orange downward facing triangles respectively. The model prediction is shown as the solid line in the same colors. The shaded areas represent the values between the 5<sup>th</sup> and 95<sup>th</sup> percentile of the uncertainty analysis. Data show three biological replicates for each point. The reactions were initiated with 9.65  $\mu\text{M}$  afEst2, 0.73  $\mu\text{M}$  mpCAR, 0.07  $\mu\text{M}$  ttPPiase, 0.43  $\mu\text{M}$  PTDH, 0.57  $\mu\text{M}$  tnPAP, 0.15  $\mu\text{M}$  tnAK, 24.7  $\mu\text{M}$  apADH, 4,000  $\mu\text{M}$  methyl 4-toluate, 500  $\mu\text{M}$  NADPH, 500  $\mu\text{M}$  NADH, 1,250  $\mu\text{M}$  ATP, 20,000  $\mu\text{M}$  phosphite, 6,000  $\mu\text{M}$  polyphosphate, 20,000  $\mu\text{M}$   $\text{MgCl}_2$ . Observed reduced  $\chi^2$  values (for original deterministic, modified deterministic, fitted quadratic and fitted linear models): 9, 5, 31, 79.

However, the optimized pathway performed significantly better in productivity compared to the non-optimized pathway.

### Sensitivity Analysis Comparison of the Optimized vs Non-Optimized Pathways

As well as delivering optimal performance at lowest cost, it is important that industrial processes are robust, and not easily affected by issues with a single reagent. Therefore, a Sobol sensitivity analysis was carried out on the model for the optimized reaction to determine the main sources of the large uncertainty in the final alcohol concentration<sup>[65]</sup> (Figure 7). This revealed that the final alcohol yield was most sensitive to the parameters for tnAK in this optimized reaction (> 30% of total uncertainty associated with these). Other parameters with a substantial contribution included polyphosphate concentration, CAR inhibition by  $\text{PP}_i$ , and the rate of aldehyde degradation. These parameters were very different from those observed in the non-optimized reaction (Figure 7). This indicated that, whilst the optimization had been highly successful in delivering greater productivity at lower cost, this came at the cost of increased sensitivity to some parameters related to individual

enzymes (resulting in a 3.5 fold increase in the 95% CIs of the final yield).

## Discussion

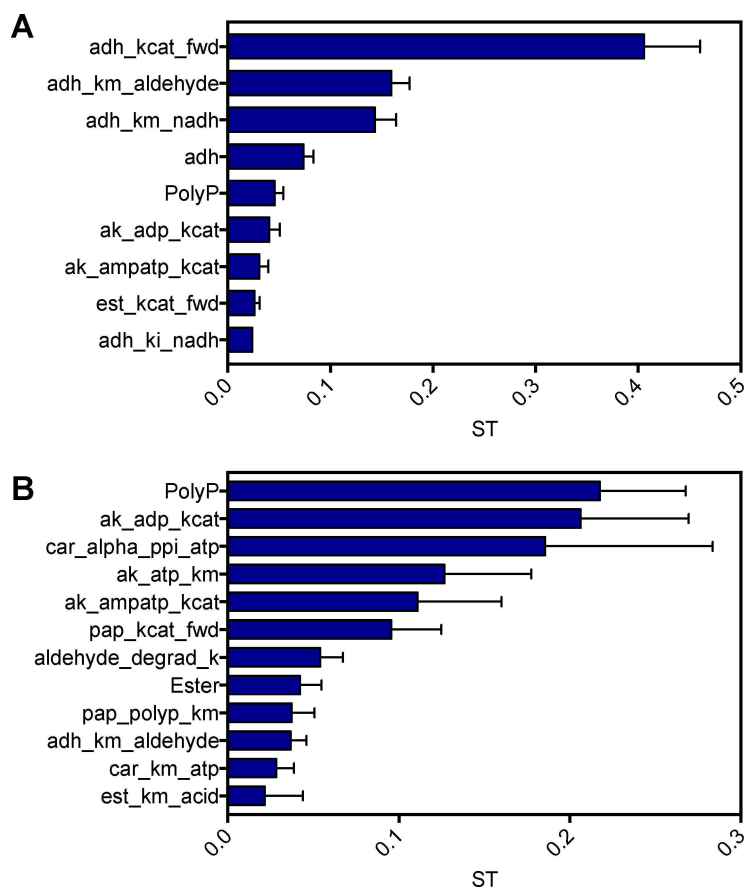
### Multi-step *in vitro* Biocatalysis

Among several other advantages, the use of isolated enzymes allows reactions to be easily controlled and subsequently optimized quickly and easily. We have demonstrated the use of CARs for multi-step *in vitro* biocatalysis, making use of mechanistic modelling to understand the dynamics of multiple enzymes working in concert. The modular nature of constructing models for each enzyme allows a systems biology approach to be taken for the creation of new cascades, with the effects of enzyme module addition or removal being predictable *in silico*. This approach requires more experimental data than models with parameters fitted to product outputs (e.g.<sup>[41]</sup>). However, the final outcome has greater predictive capability ( $\chi^2$  in Figure 5 and 6 is substantially better than simple equations fitted to the data, in contrast to models fitted to the product outputs). Furthermore, establishing such a mathematical model for this seven enzyme *in vitro* cascade facilitates its understanding. Key advantages include the ability to exploit the model for the evaluation of different process options, identification of bottlenecks and optimization of the reaction. Alternative strategies require costly systematic testing and optimization of enzyme concentrations.<sup>[66,67]</sup>

### Modelling Drives Hypothesis Generation

Most importantly, modelling was critical in identifying the inhibition of CAR by  $\text{PP}_i$  (Figure 2). This has since been independently shown by Kunjapur et al.<sup>[19]</sup> This demonstrates one benefit of mechanistic modelling over a purely data driven approach. Upon observing the slower rate of the CAR reaction compared to the model expectations, we were able to focus efforts on determining the cause of this.<sup>[15]</sup> We were then able to test the addition of the PPiase enzyme *in silico* before committing to the process of preparing this enzyme for inclusion in the cascade. It is possible that other enzymes might have unidentified inhibitors or activators in the cascade. However, our modelling showed no evidence of these, and so provides a data-led rationale for avoiding the costly testing of all possible pairwise interactions.

Testing our observations against model predictions revealed some other interesting phenomena. The inclusion of PTDH allowed the CAR reaction to proceed more effectively than the model predicted at low NADPH concentrations (Figure 3D). More detailed modelling would be required to fully understand this. However, this observation might be a result of substrate shuttling or channeling between PTDH and the reductase domain of the CAR enzyme.<sup>[68]</sup> It could be expected that the local concentration of  $\text{NADP}^+$  around the CAR enzyme might



**Figure 7.** Sensitivity analysis of the modelled pre- and post-optimization batch reaction. The total sensitivity indices (ST) are shown which take into account 1<sup>st</sup> order and all other interactions. Total of all STs for a reaction sum to 1. Sensitivity is in reference to the uncertainty in the final 4-tolyl alcohol concentration. Error bars show the 95% confidence intervals. The sum of all sensitivity indices should equal 1. Parameters with a ST of less than 0.02 are not shown. A: Sensitivity analysis of the pre-optimization complete reaction, shown in figure 5 C. B: Sensitivity analysis of the optimized complete reaction, shown in figure 6.

be higher, and that here PTDH would work more efficiently providing a larger local supply of NADPH.<sup>[69]</sup>

### The Use of CARs *in vitro*

CARs have previously been used in whole-cell biocatalysis where ATP and NADPH can be regenerated by the host metabolism.<sup>[70]</sup> Using CARs as isolated enzymes allows the design of an efficient process, and the inclusion of cofactor regeneration makes this an economically competitive alternative to whole-cell biocatalysis. PTDH accepts two substrates, NAD<sup>+</sup> and NADP<sup>+</sup>, which compete for the enzyme's active site. In our optimized reaction both NAD<sup>+</sup> and NADP<sup>+</sup> were set to 500  $\mu$ M and only PTDH concentration was optimized. In this situation NAD<sup>+</sup> is regenerated more effectively as PTDH has a lower  $K_M$  and higher  $k_{cat}$  for this substrate. Possibly the PTDH reaction could be optimized further by allowing the ratio of NAD<sup>+</sup>:NADP<sup>+</sup> to be altered in the optimization procedure, to balance NADH and NADPH regeneration to the requirements of the reaction.

### The ATP Regeneration System is Challenging to Model

The reaction catalyzed by AK is critically regulated by magnesium.<sup>[71]</sup> While AK has been shown to follow a random Bi Bi mechanism,<sup>[72]</sup> knowledge of the free and bound magnesium concentrations is required to implement this mechanism as a rate equation. This is highly challenging as the stoichiometry and dissociation constant of the magnesium-polyphosphate interaction is very poorly described. We elected to approximate the AK reaction using a bi-substrate equation (Table 2), with high levels of uncertainty associated with each parameter. The uncertainty analysis performed easily captures the potential positions of equilibrium that would be modelled using a more complex but accurate rate equation, facilitating the use of the approximation.<sup>[71]</sup> We attempted to model ATP regeneration using more realistic equations and very broad estimates of the magnesium-polyphosphate interaction (Supplementary Figures 43–46). The outcomes were very similar to those obtained with our core model. Consequently, we are confident that the approximation used is reasonable. Future iterations of the model could seek to more accurately model the free and bound magnesium concentrations.<sup>[73]</sup>

The reaction catalyzed by PAP is also difficult to model. Polyphosphate cannot be added at a saturating concentration due to chelation of magnesium, and the chain lengths and concentrations of the various polyphosphate species are difficult to determine. Since both polyphosphate chain length and concentration affect reaction kinetics, the reaction was again approximated using a bi-substrate equation with high levels of uncertainty associated with the polyphosphate concentration (Table 2).

### Reaction Optimization

Optimization of the batch reaction, used to validate the model, demonstrates its potential for exploring new process options. Optimization of the reaction resulted in just enough aEst2 to hydrolyze all of the methyl 4-toluate in the time available, and just enough ApADH to achieve 90% of the theoretical maximum of 4-tolyl alcohol at 4 hours (Figure 6). In contrast, 4-toluic acid and 4-tolualdehyde concentrations were maintained at a low but steady concentration allowing maximum productivity by all enzymes for the entire reaction. Enzyme concentrations were minimized and productivity maximized successfully.

In comparing the uncertainty in the pre-optimization (Figure 5C) and optimized (Figure 6) complete reactions, it is clear that uncertainty in the optimized reaction has increased significantly (Figure 7). Minimizing the concentration of all enzymes so that they are each close to being rate limiting has likely increased the uncertainty, as changes in a greater number of parameters can impact on the rates.

The pre-optimization reaction did show uncertainty in 4-toluic acid and 4-tolualdehyde concentrations, yet this does not translate to a high degree of uncertainty in the 4-tolyl alcohol concentration. When 4-tolualdehyde concentration is high enough above its  $K_M$  for ApADH, the main driver for 4-tolyl alcohol concentration uncertainty is the rate of the ApADH reaction (principally the  $adh\_kcat\_fwd$  parameter; Figure 7A). However when 4-tolualdehyde is maintained at the lower concentrations in the optimized reaction, this concentration has a greater impact on the rate of 4-tolyl alcohol production. Many of the parameters causing uncertainty in 4-tolualdehyde concentration are then responsible for the larger degree of uncertainty in 4-tolyl alcohol production (Figure 7B). This suggests that the sensitivity of product 4-tolyl alcohol production could be kept low by optimizing the 4-tolualdehyde concentrations to be maintained above the ApADH  $K_M$  for this intermediate. Future optimization could include an additional objective for the minimization of uncertainty in the final 4-tolyl alcohol concentration. This would allow costs to be minimized whilst also minimizing the uncertainty in the final yield. However the inclusion of two competing objective functions would make the optimization process more complex, likely requiring the generation and manual evaluation of a set of Pareto optimal solutions.<sup>[74]</sup>

We estimate the cost of our optimized reaction (Figure 6), as approximately \$0.0014 (for a 500  $\mu$ L reaction; or \$7 per gram of a 122 Da product; Supplementary Data), when performed at

a scale to benefit from bulk prices.<sup>[75]</sup> Further reductions in the cost would likely be achieved at high scales. Re-use of the enzymes (e.g. by enzyme immobilization) would significantly reduce enzyme costs; and cofactor costs could be minimized using the cheaper spent cofactors as the starting material. This would make the cascade competitive for the production of fine chemicals (e.g. perfumes or pharmaceuticals).

### Uncertainty and Sensitivity Analysis Facilitate the Use of Approximations

The use of uncertainty analysis to incorporate error or approximations into the modelling was critical in its implementation. For example, large uncertainty bounds covering an extensive range of feasible parameter values for the reverse direction of the esterase reaction were modelled. Despite the high levels of uncertainty for these parameters, the model is still able to make a good predictions, giving low  $\chi^2$  values compared to simple models. This is particularly so for the complete reaction where error in these parameters have little impact on the prediction for the final alcohol concentration (Figures 5C, 6 and 7).

Furthermore, a number of approximations were necessary while modelling the ATP regeneration for the CAR step, resulting in a large uncertainty in the CAR reaction (Figure 3E). However high uncertainty in the CAR step did not necessarily result in high uncertainty in the complete reaction (Figure 5C and Figure 7), although some parameters had a larger effect once enzyme concentrations were minimized in the optimized reaction (Figure 6 and Figure 7). Sensitivity analysis suggests a more accurate characterization of the inhibition by  $PP_i$  on the CAR enzyme, as well as a better characterization of the AK reaction, would be good targets for reducing the uncertainty in the optimized reaction in future experiments.

### Conclusions

Multi-enzyme cascades, many of which are made possible via CAR enzymes, offer the opportunity to make an extended range of biocatalytic transformations economically viable. Here we show that thorough characterization of every component of a multi-enzyme cascade allows the development of a deterministic cascade model. This deterministic model robustly predicts the behavior of the cascade *in vitro* ( $\chi^2=3.0$ ), demonstrating that such models are readily tractable for cascades of at least seven enzymes, a considerable advance on the most complex cascade previously described. We further demonstrate the utility of this model in understanding the reaction and for rapidly identifying steps in the catalysis that do not perform as expected. Finally, we exploited the model to optimize the transformation for maximum pathway flux with minimal enzyme usage. Our findings validate the use of deterministic models for *in vitro* biocatalysis, and strongly suggest that these should be used more widely in process development to increase the economic efficiency of enzyme cascades.

## Experimental Section

### Materials and Plasmids

All chemicals were purchased from Sigma-Aldrich (Gillingham, UK), and were of the highest purity available. Plasmid files are available as supplementary file 2. pNIC28-Bsa4<sup>[76]</sup> is available through Addgene. Other pET vectors used in this study are available from Novagen. All reactions were performed in aqueous solutions. Methyl 4-toluate (#259667), 4-toluic acid (#T36803), 4-tolualdehyde (#35602) and 4-tolylalcohol (#127809) were each prepared at stock concentration of 500 mM in DMSO.

### Enzyme Preparation

Genes for the expression of mpCAR, PTDH, tnAK, tnPAP, and ttPPIase were cloned into the *Nco*I and *Hind*III sites of pNIC28-Bsa4.<sup>[76]</sup> The native *apADH* gene sequence was cloned previously into pET-30 Xa/LIC.<sup>[77]</sup> The gene for mpCAR was obtained by PCR from genomic DNA. All other genes were codon optimized for *E. coli* and gene synthesized by IDT. All contained a N-terminal 6x histidine tag<sup>[76]</sup> (sequences available in Supplementary file 2). Vectors were transformed into BL21 (DE3) *E. coli* for expression. mpCAR was co-transformed with a pCDF-Duet1 vector containing a phosphopantetheine transferase from *Bacillus subtilis*. A pET24-d plasmid containing *afEst2* transformed into the *E. coli* strain BL21-CodonPlus (DE3)-RIPL (Agilent) has been described previously.<sup>[44]</sup> Expression was carried out in LB media with the addition of 50 µg/ml of appropriate antibiotics.

Cells were grown to approximately 0.6 OD<sub>600nm</sub> at 37 °C with shaking at 225 rpm, at which point IPTG was added to a concentration of 100 µM. Cells were then incubated overnight at 20 °C, except for *afEst2* which was incubated overnight at 30 °C. Cells were harvested by centrifugation at 4,750 *g* and re-suspended in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl.

Cell lysate was prepared by sonication on ice followed by centrifugation at 20,000 *g* at 4 °C for 30 min to remove the insoluble fraction. Enzymes were purified from the cell lysate using a 1 ml His-Trap FF crude column (GE Healthcare) using an elution gradient from 10 to 250 mM imidazole in 25 mM Tris-HCl pH 8.0, 0.5 M NaCl. The purified sample was then applied to a Superdex 200 HiLoad 16/60 gel filtration column (GE Healthcare) and eluted in 25 mM HEPES, pH 7.5, 0.1 M NaCl at 1.0 ml/min. Eluted fractions were analyzed by SDS-PAGE before being pooled and concentrated to between 1 and 8 mg/ml. To calculate protein concentration from OD<sub>280nm</sub>, an extinction coefficient and molecular weight for each enzyme was calculated using the ExPaSy ProtParam tool.<sup>[78]</sup> Single use aliquots of protein were stored at –80 °C.

### Enzyme Assays

Unless otherwise stated, all reactions were performed in triplicate in a 96-well microtiter plate using a Tecan M200 Infinite plate reader. All assays were carried out in a standard reaction buffer consisting of 100 mM Tris-HCl at pH 7.5, titrated to the correct pH whilst at 30 °C. All experiments used three experimental replicates, defined as experiments set up and run independently for each condition tested. Triplicate samples were sufficient to determine the necessary constants, whilst permitting sufficient throughput in single experiments.

### Buffers for pH vs Activity Assays

Buffers to measure activity across a range of pH values were prepared at 0.2 pH unit intervals at assay temperature. Buffers used were: 50 mM citrate between pH 4.0 and 5.8, 50 mM MES between pH 5.8 and 6.6, 50 mM PIPES between pH 6.4 and 7.4, 50 mM MOPS between pH 6.8 and 7.8, 50 mM HEPES between pH 7.0 and 8.0, 50 mM Tris between pH 7.8 and 9.0, and 50 mM boric acid between pH 8.8 and 10. 10 M and 1 M HCl or NaOH was used to titrate the buffers to the correct pH as appropriate.

### Thermostability Assays

Thermostability was measured by heating enzyme samples for 30 min using the temperature gradient of a Bio-Rad thermocycler between 30 and 90 °C (or less where appropriate) before cooling on ice. Activity was then measured using specific enzyme assays (below) relative to a control sample kept on ice. Where appropriate, data were adapted from previous work as indicated in the relevant sections.

### CAR Assays

To determine the two substrate kinetics of mpCAR with ATP and 4-toluic acid, reactions were set up containing 11 µg / ml mpCAR enzyme, 0.25 mM NADPH, 10 mM MgCl<sub>2</sub>, with a range of ATP and 4-toluic acid concentrations. The oxidation of NADPH was used to monitor reactions by measuring absorbance of NADPH at 340 nm. Reactions were carried out for 10 minutes after a 5 minute pre-incubation at 30 °C. The appropriate equation was determined by fitting the initial rates of reaction at eight ATP concentrations around the expected *K<sub>M</sub>* including a blank, each at five 4-toluic acid concentrations around its expected *K<sub>M</sub>*. Data were fitted by least squares non-linear regression using GraphPad Prism 5.0, and possible reaction mechanisms compared for goodness of fit. A *K<sub>M</sub>* for NADPH was determined by setting up reactions containing 10 µg/ml mpCAR enzyme, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 10 mM 4-toluic acid and a range of eight NADPH concentrations around its expected *K<sub>M</sub>*, after confirming NADPH concentration at OD<sub>340 nm</sub>. Other characterization was performed previously.<sup>[15]</sup>

### ADH Assays

To determine kinetic parameters for *apADH*, assays were carried out in sealed PCR tubes using a Bio-Rad thermocycler. After 15 to 30 minutes reactions were quenched by transfer into a 96-well microtiter plate containing 5 mM NaOH. Increase or decrease in NADH concentration was measured at OD<sub>340nm</sub> to determine activity, relative to a blank reaction. Reactions to measure activity in the reductive direction were set up containing either 0.5 mM NADH or 10 mM 4-tolualdehyde. Reactions to measure activity in the oxidative direction were set up containing either 2.5 mM NAD<sup>+</sup> or 100 mM 4-tolyl alcohol. In each case a range of concentrations were used for the other respective substrate around its expected *K<sub>M</sub>*, including a blank. 4-tolylaldehyde kinetics were carried out at 30 °C in standard reaction buffer using 166.7 µg/ml *apADH* enzyme. NAD<sup>+</sup>, NADH and 4-tolyl alcohol kinetics were carried out at 70 °C in 100 mM Tris-HCl titrated to pH 7.5 at 70 °C, using 40 µg/ml *apADH* enzyme. Only *K<sub>M</sub>* constants were taken from data at 70 °C. Data were fitted by least squares non-linear regression using GraphPad Prism 5.0. Data for the effects of pH and temperature were adapted from previous work.<sup>[48]</sup>

### PTDH Assays

To determine kinetic parameters for PTDH, assays were set up containing 20 mM Na<sub>2</sub>HPO<sub>3</sub>, 2.8 µg / ml PTDH enzyme, and a range of NAD<sup>+</sup> or NADP<sup>+</sup> concentrations around the expected  $K_M$ . The production of NADH or NADPH was monitored at OD<sub>340nm</sub> and data were fitted by least squares non-linear regression using GraphPad Prism 5.0. The thermostability of a sample containing 18 µg / ml PTDH was measured as described above. Activity at various pH values was determined using a range of buffers described previously, in place of the standard reaction buffer.

### PPiase

PPiase activity was measured using the production of phosphomolybdate to measure phosphate content, as described.<sup>[79]</sup> Using a standard curve, the rate of phosphate production by ttPPiase at 0.22 µg/ml was determined using a high (5 mM) concentration of PP<sub>i</sub> in the presence of 10 mM MgCl<sub>2</sub>, in standard reaction buffer. Five measurements were taken in triplicate over 20 minutes. A conservative  $K_M$  was estimated from the BRENDA database<sup>[55]</sup> and the rate of PP<sub>i</sub> hydrolysis used to calculate  $k_{cat}$ .

Activity at variable pH was determined using the same phosphomolybdate method with reaction buffers covering a range of pH values, detailed above. The thermostability of a sample containing 0.44 µg/ml ttPPiase was measured using the assay described previously, with readings taken at time points 0 and 20 minutes. 10 mM MgCl<sub>2</sub> was included in the heated sample.

### PAP Assays

ADP formation was measured using an ADP-Glo kinase assay kit from Promega in a 384-well solid white microtiter plate, following the manufacturer's instructions. Assays were carried out in standard reaction buffer containing 20 or 40 mM MgCl<sub>2</sub>, 12.5 mM polyphosphate, 3.25 mM AMP, 12.5 µg/ml or 125 µg/ml of tnPAP enzyme (as indicated in SI). Polyphosphate was calculated as concentration of Na<sub>3</sub>PO<sub>3</sub> units, with a molecular weight of 102. Reactions were carried out for 15 minutes before quenching with the ADP-Glo kit. A range of polyphosphate, AMP and MgCl<sub>2</sub> concentrations around an expected  $K_M$  were assayed in turn. ADP production was calculated from a standard curve, and the data fitted by least squares non-linear regression using GraphPad Prism 5.0. Specific assay conditions are shown in Supplementary Figures 35 to 38. Activity at different pH values was measured using a range of buffers described above in place of the standard reaction buffer. Thermostability was determined as described above, using the ADP-Glo assay to measure residual relative activity.

### Mathematical Modelling

Mathematical modelling was carried out in Python 3.4 using the SciPy<sup>[80]</sup> and SALib<sup>[81]</sup> modules. A python package (<https://github.com/willfinnigan/kinetics>) was developed during the course of this work. The models created in this study are available at [https://github.com/willfinnigan/Seven\\_enzyme\\_biotransformation\\_2019](https://github.com/willfinnigan/Seven_enzyme_biotransformation_2019). Outputs were exported to GraphPad Prism 5.0 for plotting. The integrate.odeint function in SciPy was used to solve ordinary differential equations.

Uncertainty analysis was carried out making use of the SALib module.<sup>[81]</sup> Input bounds for each parameter were defined as either the calculated 95% confidence intervals or where uncertainty was judged to be high 25 or 50% of the parameter value. Sampling of the possible inputs within these bounds was carried out by Latin

hypercube sampling of 1,000 samples, with the model for each sample run. At each time point the mean, 5<sup>th</sup> and 95<sup>th</sup> percentile of each analyte concentration was plotted to represent the uncertainty of the model.

Reduced chi-squared values were calculated using Equation (20):<sup>[51]</sup>

$$\chi_{red}^2 = \frac{1}{K} \sum_{n=1}^N \left( \frac{|y_n| - f(x_n)}{\frac{\sigma_n}{\sum_{m=1}^M |y_m|}} \right)^2 \quad (20)$$

Where  $K$  is the number of degrees of freedom;  $N$  is the total number of experimental points determined;  $M$  is the total number of experimental points determined for each substrate/intermediate/product;  $|y_{m/n}|$  is the mean experimental value at each point determined;  $f(x_n)$  is the model derived value for  $y$  at time  $x$ ; and  $\sigma_n$  is the standard deviation in each experimental value. The reduced chi-squared formula was modified here to take the mean fractional standard deviation for each chemical: this approach was taken to mitigate the effect of (random) tightly determined triplicate points dominating the calculation. Three models were used for each dataset: the median deterministic model described above; a quadratic equation of the form  $y = ax^2 + bx + c$  fitted to the experimental data; and a linear equation of the form  $y = ax + c$  fitted to the experimental data.

The method of Sobolj<sup>[82]</sup> was used for sensitivity analysis as part of the SALib module,<sup>[81]</sup> although numerous methods are available. Second order effects were not calculated, and the total sensitivity indices plotted to show the sensitivity of each parameter. The sample number was set at 1,000 (chosen for consistency with past studies; this was sufficient to obtain a good sampling rate whilst not overloading computing resources). For more information on the use of uncertainty and sensitivity analysis in process modelling please see the following references.<sup>[83,84]</sup>

### Optimization Using a Genetic Algorithm

A custom built, single objective genetic algorithm was used to minimize enzyme cost, on the condition over 90% alcohol yield was reached. A general outline of the algorithm is shown in Supplementary Figure 42.

### Model Validation Reactions

Standard curves of methyl 4-toluate, 4-toluic acid, 4-tolylaldehyde and 4-tolyl alcohol were prepared and analyzed by HPLC. 500 µl reactions were set up in triplicate in 1.5 ml microcentrifuge tubes, and incubated in a thermoshaker (EpMotion T5075 t thermal module, Eppendorf) at 30 °C with 500 rpm shaking. 50 µl samples were taken every 30 minutes, mixed with 50 µl acetonitrile, and centrifuged for 10 minutes. Supernatant was removed and stored at 4 °C before being analyzed by HPLC.

### HPLC

An Eclipse Plus C18 column (Agilent) with a particle size of 3.5 µm, measuring 4.6 x 100 mm, was used. The column was run at 60 °C on the following method using two eluants: eluant A: 95% H<sub>2</sub>O, 5% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid; eluant B: 5% H<sub>2</sub>O, 95% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid. 3 µl of sample was injected and eluted on a gradient from 0 to 100% eluant B over 10 minutes. Eluant B was maintained at 100% for a further 2 minutes before the column was re-equilibrated with eluant A for 2 minutes.

## Mass Spectrometry

Low and high resolution mass spectra were obtained by staff at the University of Manchester. Electrospray (ES) spectra were recorded on a Waters Platform II with an SQ Detector 2. High resolution mass spectra (HRMS) were recorded on a Thermo Finnigan MAT95XP and are accurate to  $\pm 0.001$  Da.

## Acknowledgements

The authors thank Andrew Hill (University of Manchester) for providing the pCDF-Sfp plasmid. We also thank Colin Edge, Dan Tray and Neil Hodnett (GlaxoSmithKline) for their advice on process modelling; and Dr Khulood Alyahya (University of Exeter) for advice on model quality. Additionally we thank Anna Peters (University of Manchester) for her assistance with the mass spectrometry. WF was funded by BBSRC (grant no. BB/K501001/1) and GlaxoSmithKline. RC was funded by BBSRC and GSK.

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** Enzyme cascade · Kinetics · Pathway optimization · Isolated enzymes · Reaction engineering

- [1] D. J. Pollard, J. M. Woodley, *Trends Biotechnol.* **2007**, *25*, 66–73.
- [2] N. J. Turner, M. D. Truppo, *Curr. Opin. Chem. Biol.* **2013**, *17*, 212–214.
- [3] I. Oroz-Guinea, E. García-Junceda, *Curr. Opin. Chem. Biol.* **2013**, *17*, 236–249.
- [4] A. Wells, H. P. Meyer, *ChemCatChem* **2014**, *6*, 918–920.
- [5] K. M. Wilding, S. M. Schinn, E. A. Long, B. C. Bundy, *Curr. Opin. Biotechnol.* **2018**, *53*, 115–121.
- [6] R. du Preez, K. G. Clarke, L. H. Callanan, S. G. Burton, *J. Mol. Catal. B* **2015**, *119*, 48–53.
- [7] M. Bujara, M. Schümperli, S. Billerbeck, M. Heinemann, S. Panke, *Biotechnol. Bioeng.* **2010**, *106*, 376–389.
- [8] C. Hold, S. Panke, *J. R. Soc. Interface* **2009**, *6*, S507–S521.
- [9] R. Xue, J. M. Woodley, *Bioresour. Technol.* **2012**, *115*, 183–195.
- [10] M. Winkler, *Curr. Opin. Chem. Biol.* **2018**, *43*, 23–29.
- [11] G. Qu, J. Guo, D. Yang, Z. Sun, *Green Chem.* **2018**, *20*, 777–792.
- [12] N. J. Turner, E. O'Reilly, *Nat. Chem. Biol.* **2013**, *9*, 285–288.
- [13] S. P. France, S. Hussain, A. M. Hill, L. J. Hepworth, R. M. Howard, K. R. Mulholland, S. L. Flitsch, N. J. Turner, *ACS Catal.* **2016**, *6*, 3753–3759.
- [14] A. N. Khusnutdinova, R. Flick, A. Popovic, G. Brown, A. Tchigvintsev, B. Nocek, K. Correia, J. C. Joo, R. Mahadevan, A. F. Yakunin, *Biotechnol. J.* **2017**, *12*, 1–12.
- [15] W. Finnigan, A. Thomas, H. Cromar, B. Gough, R. Snajdrova, J. P. Adams, J. A. Littlechild, N. J. Harmer, *ChemCatChem* **2017**, *9*, 1005–1017.
- [16] E. H. Hansen, B. L. Møller, G. R. Kock, C. M. Bünner, C. Kristensen, O. R. Jensen, F. T. Okkels, C. E. Olsen, M. S. Motawia, J. Hansen, *Appl. Environ. Microbiol.* **2009**, *75*, 2765–2774.
- [17] P. Kallio, A. Pásztor, K. Thiel, M. K. Akhtar, P. R. Jones, *Nat. Commun.* **2014**, *5*, 4731.
- [18] Y.-H. P. Zhang, *Biotechnol. Adv.* **2015**, *33*, 1467–1483.
- [19] A. M. Kunjapur, B. Cervantes, K. L. J. Prather, *Biochem. Eng. J.* **2016**, *109*, 19–27.
- [20] Y. Ni, D. Holtmann, F. Hollmann, *ChemCatChem* **2014**, *6*, 930–943.
- [21] R. Woodyer, W. A. Van der Donk, H. Zhao, *Biochemistry* **2003**, *42*, 11604–11614.
- [22] M. J. McLachlan, T. W. Johannes, H. Zhao, *Biotechnol. Bioeng.* **2007**, *99*, 268–274.
- [23] H. Zhao, W. A. Van Der Donk, *Curr. Opin. Biotechnol.* **2003**, *14*, 583–589.
- [24] R. D. Woodyer, T. W. Johannes, H. Zhao, in *Enzym. Technol.* Asiatech Publishers Inc., New Delhi, India, **2003**, pp. 85–103.
- [25] A. Kameda, T. Shiba, Y. Kawazoe, Y. Satoh, Y. Ihara, M. Munekata, K. Ishige, T. Noguchi, *J. Biosci. Bioeng.* **2001**, *91*, 557–563.
- [26] G. A. Strohmeier, I. C. Eiteljörg, A. Schwarz, M. Winkler, *Chem. Eur. J.* **2019**, *6119*–6123.
- [27] S. M. Resnick, A. J. B. Zehnder, *Appl. Environ. Microbiol.* **2000**, *66*, 2045–2051.
- [28] J. I. Ramsden, R. S. Heath, S. R. Derrington, S. L. Montgomery, J. Mangas-Sanchez, K. R. Mulholland, N. J. Turner, *J. Am. Chem. Soc.* **2019**, *141*, 1201–1206.
- [29] M. H. Medema, R. van Raaphorst, E. Takano, R. Breitling, *Nat. Rev. Microbiol.* **2012**, *10*, 191–202.
- [30] D. Kuhn, L. M. Blank, A. Schmid, B. Bühler, *Eng. Life Sci.* **2010**, *10*, 384–397.
- [31] T. Ellis, T. Adie, G. S. Baldwin, *Integr. Biol. (Camb.)* **2011**, *3*, 109–18.
- [32] O. Purcell, B. Jain, J. R. Karr, M. W. Covert, T. K. Lu, *Chaos* **2013**, *23*, 025112.
- [33] D. Vasić-Rački, Z. Findrik, A. Vrsalović Presečki, *Appl. Microbiol. Biotechnol.* **2011**, *91*, 845–856.
- [34] R. H. Ringborg, J. M. Woodley, *React. Chem. Eng.* **2016**, *1*, 10–22.
- [35] G. Sin, J. M. Woodley, K. V. Gernaey, *Biotechnol. Prog.* **2009**, *25*, 1529–1538.
- [36] L. Rios-Solis, P. Morris, C. Grant, A. O. O. Odeleye, H. C. Hailes, J. M. Ward, P. A. Dalby, F. Baganz, G. J. Lye, *Chem. Eng. Sci.* **2015**, *122*, 360–372.
- [37] S. Peri, S. Karra, Y. Y. Lee, M. N. Karim, *Biotechnol. Prog.* **2007**, *23*, 626–637.
- [38] Z. Findrik, D. Vasić-Rački, *Biotechnol. Bioeng.* **2007**, *98*, 956.
- [39] G. D. Yadav, K. Manjula Devi, *JAOCs, J. Am. Oil Chem. Soc.* **2001**, *78*, 347–351.
- [40] N. A. Sayar, B. H. Chen, G. J. Lye, J. M. Woodley, *Biochem. Eng. J.* **2009**, *47*, 10–18.
- [41] J. A. Rollin, J. Martin del Campo, S. Myung, F. Sun, C. You, A. Bakovic, R. Castro, S. K. Chandrayan, C.-H. Wu, M. W. W. Adams, *Proc. Mont. Acad. Sci.* **2015**, *112*, 4964–4969.
- [42] J. H. Schrittwieser, J. Sattler, V. Resch, F. G. Mutti, W. Kroutil, *Curr. Opin. Chem. Biol.* **2011**, *15*, 249–256.
- [43] J. E. Guy, M. N. Isupov, J. A. Littlechild, *J. Mol. Biol.* **2003**, *2836*, 1041–1051.
- [44] C. Sayer, W. Finnigan, M. N. Isupov, M. Levisson, S. W. M. Kengen, J. van der Oost, N. J. Harmer, J. A. Littlechild, *Sci. Rep.* **2016**, *6*, 25542.
- [45] C. F. C. Bonting, G. J. J. Kortstee, A. J. B. Zehnder, *J. Bacteriol.* **1991**, *173*, 6484–6488.
- [46] M. Sato, Y. Masuda, K. Kirimura, K. Kino, *J. Biosci. Bioeng.* **2007**, *103*, 179–84.
- [47] C. Vieille, H. Krishnamurthy, H.-H. Hyun, A. Savchenko, H. Yan, J. G. Zeikus, *Biochem. J.* **2003**, *372*, 577–85.
- [48] H. Hirakawa, N. Kamiya, Y. Kawarabayashi, T. Nagamune, *J. Biosci. Bioeng.* **2004**, *97*, 202–206.
- [49] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics, 4th Edition*, Wiley, **2012**.
- [50] N. J. Harmer, M. Vivoli Vega, in *Biomol. Bioanal. Tech.* (Ed.: V. Ramesh), **2019**, pp. 179–217.
- [51] R. Andrae, T. Schulze-Hartung, P. Melchior, *arXiv*, **2010**, 1012.3754.
- [52] J. Kraut, *Annu. Rev. Biochem.* **1977**, *46*, 331–358.
- [53] V. Schellenberger, H.-D. Jakubke, *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1437–1449.
- [54] H. Gaertner, A. Pugserver, *Eur. J. Biochem.* **1989**, *181*, 207–213.
- [55] I. Schomburg, A. Chang, S. Placzek, C. Sohngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, *Nucleic Acids Res.* **2013**, *41*, 764–772.
- [56] H. Itoh, T. Shiba, *J. Bacteriol.* **2004**, *186*, 5178–5181.
- [57] V. Perrier, S. Burlacu-Miron, S. Bourgeois, W. K. Surewicz, A. M. Gilles, *J. Biol. Chem.* **1998**, *273*, 19097–19101.
- [58] D. Gahloth, M. S. Dunstan, D. Quaglia, E. Klumbys, M. P. Lockhart-Cairns, A. M. Hill, S. R. Derrington, N. S. Scrutton, N. J. Turner, D. Leys, *Nat. Chem. Biol.* **2017**, *13*, 975–981.
- [59] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, **2004**.
- [60] H. Mu, S. M. Zhou, Y. Xia, H. Zou, F. Meng, Y. Bin Yan, *Int. J. Mol. Sci.* **2009**, *10*, 2849–2859.
- [61] A. Mitic, T. Skov, K. V. Gernaey, *Green Process. Synth.* **2017**, *6*, 353–361.
- [62] W. A. Bubbs, H. A. Berthon, P. W. Kuchel, *Bioorg. Chem.* **1995**, *23*, 119–130.
- [63] J. L. Lin, L. Palomec, I. Wheeldon, *ACS Catal.* **2014**, *4*, 505–511.

- [64] I. Wheeldon, S. D. Minter, S. Banta, S. C. Barton, P. Atanassov, M. Sigman, *Nat. Chem.* **2016**, *8*, 299–309.
- [65] J. Nossent, P. Elsen, W. Bauwens, *Environ. Model. Softw.* **2011**, *26*, 1515–1525.
- [66] J. M. Sperl, V. Sieber, *ACS Catal.* **2018**, *8*, 2385–2396.
- [67] X. Yu, T. Liu, F. Zhu, C. Khosla, *Proc. Mont. Acad. Sci.* **2011**, *108*, 18643–18648.
- [68] D. Tullman-Ercek, *Nat. Chem. Biol.* **2015**, *11*, 180–181.
- [69] M. Zheng, S. Zhang, G. Ma, P. Wang, *J. Biotechnol.* **2011**, *154*, 274–280.
- [70] M. K. Akhtar, N. J. Turner, P. R. Jones, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 87–92.
- [71] J. M. D. Blair, *Eur. J. Biochem.* **1970**, *13*, 384–390.
- [72] X. R. Sheng, X. Li, X. M. Pan, *J. Biol. Chem.* **1999**, *274*, 22238–22242.
- [73] K. C. Vinnakota, F. Wu, M. J. Kushmerick, D. A. Beard, *Chapter 2 Multiple Ion Binding Equilibria, Reaction Kinetics, and Thermodynamics in Dynamic Models of Biochemical Pathways*, Elsevier Inc., **2009**.
- [74] K. Deb, S. Agrawal, A. Pratab, T. Meyarivan, *Proc. 6{th} Conf. Parallel Probl. Solving from Nat.* **2000**, 849–858.
- [75] P. Tufvesson, J. Lima-Ramos, M. Nordblad, J. M. Woodley, *Org. Process Res. Dev.* **2011**, *15*, 266–274.
- [76] P. Savitsky, J. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-brown, O. Gileadi, *J. Struct. Biol.* **2010**, *172*, 3–13.
- [77] A. M. Hickey, *Unpublished Work*, **2008**.
- [78] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. . Wilkins, A. R.D, A. Bairoch, "Protein Identification and Analysis Tools on the ExPASy Server," **2005**.
- [79] J. K. Heinonen, R. J. Lahti, *Anal. Biochem.* **1981**, *113*, 313–317.
- [80] E. Jones, T. Oliphant, P. Peterson, "SciPy: Open source scientific tools for Python," **2001**, <http://www.scipy.org/>.
- [81] J. Herman, W. Usher, *J. Open Source Softw.* **2017**, *2*, 10–11.
- [82] I. M. Sobol, *Math. Comput. Simul.* **2001**, *55*, 271–280.
- [83] G. Sin, K. V Gernaey, A. E. Lantz, *Biotechnol. Prog.* **2009**, *25*, 1043–1053.
- [84] J. A. Price, N. Mathias, J. Woodley, J. Kjobsted, *Proc. 12th IFAC Symp. Comput. Appl. Biotechnol.* **2013**, *12*, 161–168.

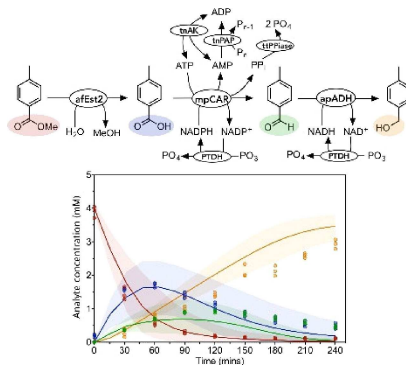
---

Manuscript received: April 9, 2019  
 Revised manuscript received: May 30, 2019  
 Accepted manuscript online: June 5, 2019  
 Version of record online: ■■■, ■■■■



## FULL PAPERS

**What a (bio)transformation!** We demonstrate the construction and validation of a kinetic model for a seven enzyme, three step biotransformation (including cofactor regeneration) using isolated enzymes. We employed the model to optimize the entire reaction using a genetic algorithm. The validated reaction demonstrates that complex multi-step reactions can be reliably modelled and optimized.



*Dr. W. Finnigan, R. Cutlan, Dr. R. Snajdrova, Dr. J. P. Adams, Prof. J. A. Littlechild, Prof. N. J. Harmer\**

1 – 17

**Engineering a Seven Enzyme Biotransformation using Mathematical Modelling and Characterized Enzyme Parts**

