Heterogeneous & Homogeneous & Bio- & Nano-
CHEMCATCHEM
CATALYSIS

# Accepted Article

**Title:** Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry

**Authors:** William Finnigan, Adam Thomas, Holly Cromar, Ben Gough, Radka Snajdrova, Joseph P Adams, Jennifer A Littlechild, and Nicholas Harmer



**To be cited as:** *ChemCatChem* 10.1002/cctc.201601249

**Link to VoR:** http://dx.doi.org/10.1002/cctc.201601249

A Journal of
ChemPubSoc Europe

WILEY-VCH www.chemcatchem.org

10.1002/cctc.201601249

WILEY-VCH

FULL PAPER

# Characterization of carboxylic acid reductases as enzymes in the toolbox for synthetic chemistry

William Finnigan [a], Adam Thomas[a], Holly Cromar[a], Ben Gough[a], Radka Snajdrova[b], Joseph P. Adams[b], Jennifer A. Littlechild[a], Nicholas J. Harmer*[a]

**Abstract:** Carboxylic acid reductase enzymes (CARs) meet the demand in synthetic chemistry for a green and regio-specific route to aldehydes from their respective carboxylic acids. However, relatively few of these enzymes have been characterized. A sequence alignment with members of the ANL superfamily of enzymes shed light on CAR functional dynamics. Using a phylogenetic analysis of known and hypothetical CARs, four unstudied enzymes were selected, and for the first time, a thorough biochemical characterization carried out. Kinetic analysis of these enzymes with various substrates shows they have a broad, but similar substrate specificity. Electron rich acids are favored, suggesting that the first step in the proposed reaction mechanism, attack by the carboxylate on the α-phosphate of ATP, is the step determining substrate specificity and reaction kinetics. The effects of pH and temperature provide a clear operational window for the use of these CARs, while investigation of product inhibition by $NADP^+$, AMP and pyrophosphate ($PP_i$) indicates that binding of substrates at the adenylation domain is ordered with ATP binding first. This paper consolidates CARs as important and exciting enzymes in the toolbox for sustainable chemistry, providing specifications for their use as a biocatalyst.

## Introduction

The demand for 'green chemistry' is of increasing global importance, driven by the need to balance sustainable and efficient resource utilization with the demands and increasing consumption of a rising population.[1] Biological solutions to chemistry challenges are a critical component in meeting this demand. The use of isolated enzymes and cell-based systems that produce negligible dangerous waste, often with higher yields, offers an alternative to traditional chemical processes. In some cases, biological alternatives are more rapid and cost effective than their chemical counterpart.[2] Despite these potential advantages, enzymes are still under-used in chemistry. Expanding the toolbox of available enzymes is essential for the successful development of new synthetic routes and sustainable manufacturing processes.[3]

An important opportunity that is ripe for exploitation is synthetic routes based on organic acids. These compounds have a long history of production by fermentation.[4] Indeed, multiple carboxylic acids were identified to be “Top Value Added Chemicals From Biomass”,[5] many of which are now being produced industrially. Reduced products of these organic acids, especially optically pure aldehydes and alcohols, are essential building blocks for use in the chemical, pharmaceutical and food industries.[6] However, chemical methods for the reduction of carboxylic acids are limited, and require chemicals such as lithium aluminium hydride and sodium borohydride in stoichiometric amounts.[7]

Two enzyme classes are capable of reduction of organic acids to aldehydes, and a review of the biocatalytic reductions possible by organisms harboring them published.[7] The aldehyde oxidoreductases (AORs) reversibly oxidize organic aldehydes to their respective acids. The oxidized product is more thermodynamically favorable, and so the equilibrium tends towards this product. AORs are therefore more useful for syntheses that require the oxidation of aldehydes.[8,9] In contrast, the carboxylic acid reductases (CARs) catalyze the reduction of a carboxylic acid to an aldehyde at the expense of ATP and NADPH producing AMP, $PP_i$ and $NADP^+$ as by-products.[6] The reduction of carboxylic acids into aldehydes using CARs has previously been confirmed by a number of studies, using GC-MS analysis. Products other than the aldehyde have not been detected.[10,11] The hydrolysis of ATP makes the reduction of acids to aldehydes by CARs strongly thermodynamically favorable, making their use an attractive “green chemical“ route to aldehyde production.[7] This synthesis can be coupled to other enzymes such as an alcohol dehydrogenase which can provide a complete route to the alcohol product.[10]

Indeed, CARs have been employed in a number of synthetic pathways. These include the production of the flavor vanillin by yeast,[12] and a synthetic pathway for the production of propane in *Escherichia coli*.[13] These examples both highlight the potential of CARs as part of a toolbox for synthesis of fine chemicals from non-oil-based chemical precursors.[5]

CARs are relatively large, multidomain enzymes of around 130 kDa. They feature an N-terminal adenylation domain, a C-terminal thioester reductase domain that likely adopts a Rossmann fold, and a central phosphopantetheine binding domain (Figure 1).[14] A phosphopantetheine arm must be

[a] W. Finnigan, A. Thomas, H. Cromar, B. Gough, J.A. Littlechild, N. J. Harmer
Biosciences, College of Life and Environmental Sciences
University of Exeter
Stocker Road, Devon, Exeter EX4 4QD, UK
E-mail: N.J.Harmer@exeter.ac.uk

[b] R. Snajdrova, J. Adams
Synthetic Chemistry
AC - API Chem – UK
GlaxoSmithKline R&D Ltd
Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

Supporting information for this article is given via a link at the end of the document.

covalently attached to a conserved serine in this central domain through the action of a phosphopantetheine transferase for the production of an active enzyme.[15] Fungal α-aminoadipate reductases, which are responsible for the reduction of α-aminoadipate to α-aminoadipate semialdehyde in lysine biosynthesis, share this domain architecture, also with the requirement for the loading of a central phosphopantetheine prosthetic group.[16] However these enzymes have been shown to have a different substrate specificity from CARs.[11]

Phosphopantetheine arms are most commonly associated with acyl carrier proteins where they maintain an acyl chain in a energetically active thioester bond, with the length and flexibility of the arm allowing access to spatially distinct active sites.[17] In CARs, the phosphopantetheine arm is believed to act in much the same way, shuttling an attached acyl chain between the N- and C-terminal domains. [15]

The proposed mechanism of CAR enzymes has four main steps (labelled 1 to 4; Figure 1). In the first two steps, the relatively unreactive carboxylic acid is activated to form a thioester with the phosphopantetheine arm at the N-terminal adenylation domain, in a mechanism possibly similar to that of the ANL superfamily of adenylating enzymes such as long chain fatty acid CoA ligases.[18],[19] (1) ATP and a carboxylic acid enter the active site of the adenylation domain where the α-phosphate of ATP is attacked by an oxygen from the carboxylic acid, forming an AMP-acyl phosphoester with the release of pyrophosphate.[19] (2) The thiol group of the phosphopantetheine arm can then nucleophilically attack the carbonyl carbon of the AMP-acyl phosphoester intermediate, releasing AMP and forming an acyl thioester with the phosphopantetheine arm. (3) The phosphopantetheine arm transfers to the C-terminal reductase domain (4) where the thioester is reduced by NADPH, releasing the aldehyde and $NADP^+$, and regenerating the thiol of the phosphopantetheine arm in the process.[7]

Relatively few CARs have been explored to date. CARs were first described in *Neurospora crassa* as an aryl-aldehyde: $NADP^+$ oxidoreductase.[20] Recently this CAR has also been further characterized.[21] Subsequently, CARs were characterized from *Nocardia asteroides* JCM 3016 [22] and later *Nocardia iowensis* [23] (referred to as niCAR here) when they were reclassified as carboxylic acid reductases.[23] Characterization of the *Nocardia asteroides* JCM 3016 CAR was performed by comparing the relative activity of this enzyme towards various aromatic substrates (1 mM concentration). This CAR was reported to prefer 3-substituted benzoates and aliphatic acids that were substituted with a phenyl group. No reaction of this CAR with simple aliphatic acids was reported. The optimum pH for activity of this enzyme was pH 7.5, and the optimum temperature for activity was 40 °C.[22]

The relative activities of the *Nocardia iowensis* CAR (niCAR) against various aromatic substrates have also been reported. The highest activity was achieved with indole-5-carboxylic acid, which was the most activated carboxylic acid tested. Substrates with 2-substituted benzoates or ring-deactivating groups showed no or very low levels of activity. The reduction of racemic ibuprofen by whole *Nocardia iowensis* cells gave a enantiometic excess (ee) of 61.2 %, which has been attributed to enantio-selectivity by niCAR based on kinetic data for its reduction of (*S*)-(+)-Ibuprofen and (R)-(-)-Ibuprofen enantiomers [23]. The requirement for the presence of a phosphopantetheine transferase for the loading of a phosphopantetheine group onto the CAR enzyme was shown for niCAR [23] and is presumed to be the case for all the CAR enzymes.

A CAR from *Mycobacterium marinum* has also been described and its application for the reduction of fatty acids to fatty alcohols explored. This CAR is active against fatty acids between two and

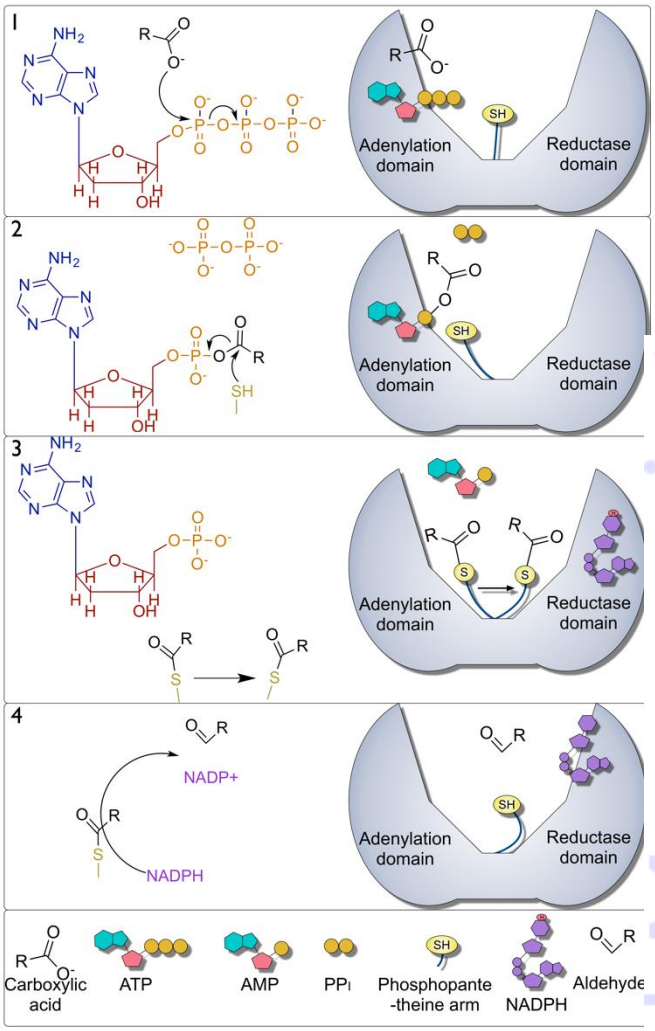


**Figure 1 – Proposed mechanism of CAR enzymes.** **1:** ATP and a carboxylic acid enter the adenylation domain where a phosphoester intermediate is formed releasing pyrophosphate in the process. **2:** the thiol of the phosphopantetheine arm nucleophilically attacks the carbonyl carbon of this intermediate forming a thioester intermediate with the phosphopantetheine arm, releasing AMP. **3:** the phosphopantetheine arm transfers to the reduction domain where, **4:** the thioester bond is reduced by NADPH releasing an aldehyde product, regenerating the phosphopantetheine thiol group and producing $NADP^+$.

eighteen carbons in length.[10] CARs have also been reported in the fungi *Syncephalastrum racemosum*[24] and *Trametes versicolor*.[25]

Recently a characterization of CARs from *Nocardia iowensis, Nocardia brasiliensis, Mycobacterium marinum* and *Mycobacterium smegmatis* showed CARs to prefer substrates where the carboxylic acid was the only polar or charged group, giving a useful insight into the substrate specificity of these enzymes. Also, a model was developed for the prediction of CAR reactivity using this and previous CAR data.[11] It is worth noting that the CAR characterized by Moura et al. is distinct from msCAR used in this study.

Here, we have produced a detailed phylogeny of the CARs and identified four previously undescribed CARs for further study that are broadly spread across this phylogeny. With the addition of niCAR for comparison to earlier work, a thorough biochemical characterization was carried out on each. We investigated the effects of temperature and pH to identify suitable conditions for the use of CARs in biocatalytic reactions. We further performed a full kinetic analysis on a range of aromatic and aliphatic substrates with these CARs to look for potential differences in their substrate specificity and to examine the effects of various functional groups on their kinetic parameters. Finally, we describe potential issues of product inhibition with the CAR enzymes. Our investigation provides a more thorough description of the factors to be considered when using the CAR enzymes in industrial biocatalysis.

**Table 1 – Carboxylic acid reductases chosen for this study.**

| Abbreviation | Source | NCBI Reference: |
|---|---|---|
| mpCAR | *Mycobacterium phlei* | WP_003889896.1 |
| msCAR | *Mycobacterium smegmatis* | AFP42026.1 |
| niCAR | *Nocardia iowensis* | Q6RKB1.1 |
| noCAR | *Nocardia otitidiscaviarum* | WP_029928026.1 |
| tpCAR | *Tsukamurella paurometabola* | WP_013126039.1 |

Five carboxylic acid reductases were chosen for a thorough biochemical characterization from a range of host organisms containing putative CAR's. niCAR has previously been characterized and was chosen for comparison.[11, 23] CAR abbreviations have been chosen to reflect their source. NCBI ascension numbers are shown which may be used to access the protein sequences. A table of the sequence identities of these five orthologues is provided as Supplementary Table 1.

## Results

### Alignment and phylogenetic analysis

CAR adenylation domains were aligned with a firefly luciferase, a fatty acyl-CoA ligase and a reductase domain from a non-ribosomal peptide synthetase, all from the ANL superfamily (Supplementary Figure 1). CARs share ~20% sequence homology with other ANL superfamily members. Members of the ANL superfamily catalyze the initial adenylation of a carboxylic acid to form an acyl-AMP intermediate, which is generally followed by the formation of a thioester. The family name is based on three of its sub families: Acyl-CoA synthetases, the non-ribosomal peptide synthase (NRPS) adenylation domain, and the Luciferase enzyme. [18] Previous alignment and crystallography studies have identified ten motifs that are conserved within the superfamily. Of the ten, five are strongly conserved within the CARs, including the active site ppxTSGSTGxPK, rGxTE and TGD motifs (where p=aliphatic and r=aromatic residues). These motifs are considered "signature" to the ANL superfamily, and are involved in the hydrolysis of ATP. [26] The remaining five motifs are also present albeit with lower conservation.

A total of 48 unique sequences showing homology to known CAR proteins were gathered using pBLAST, or mined directly from GenBank by raw text searches (Supplementary Figure 2). All sequences identified were solely from Subclass *actinobacteridae*. Within this Subclass, sequences were obtained from families *Streptomycetaceae* and *Corynebacterinae*.

A masked multiple sequence alignment of the dataset was shown to be best fit to the Whelan and Goldman model of amino acid substitution, with a discrete gamma distribution of mutation rates and an assumed presence of invariant sites (WAG+I+G). This model was implemented into a Bayesian phylogenetic reconstruction (Figure 2). According to 16S data, the *Streptomycetaceae* are thought to have evolved before the *Corynebacterinae*. However, rooting the tree on the streptomycetes has poor parsimony as numerous gene loss events would have had to have occurred for this to be the case.[27] Instead, due to an outgroup being unobtainable for this dataset we opted to root the tree on its midpoint. The tree is extremely well supported, with all nodes possessing a confidence score of >0.75, and only four of forty-six biologically relevant nodes being scored at below the highest possible confidence score of 1.

In order to better understand how CAR functionality differs across clades, we selected sequences for characterization from a range of host organisms that broadly cover distinct areas of the phylogenetic tree.

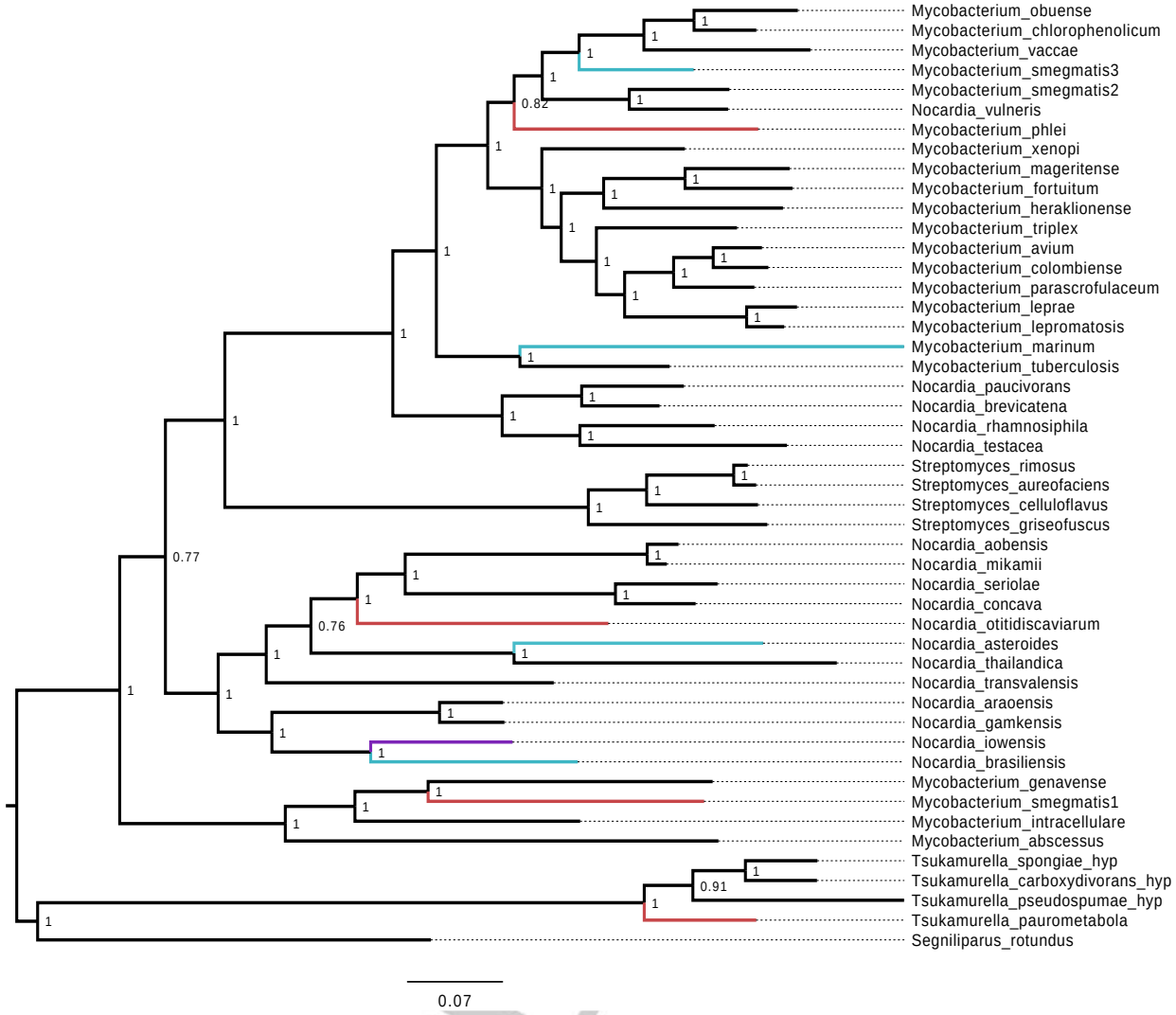


**Figure 2 – Phylogenetic tree of CAR enzymes.** A midpoint rooted phylogeny of a masked alignment of 48 carboxylic acid reductase sequences retrieved GenBank. Phylogeny was constructed with MrBayes and visualized in FigTree. Node labels represent Bayesian posterior probabilities describing node relia (with 1 being unequivocal) computed by MrBayes. Coloured branches represent CARs that have been studied: Blue – in previous research, Red – in this pa Purple – in both this paper and previous research.

### Expression and Purification

CAR enzymes (Table 1) were expressed in *E. coli* and purified from the cell lysates by nickel affinity chromatography followed by gel filtration in order to obtain a high level of purity (Supplementary Figure 3). The optimum conditions for the expression of *Mycobacterium phlei* CAR (mpCAR) in *E. coli* in LB media were determined to be induction at $OD_{600}$ 0.6 with 150 μM IPTG, followed by incubation for approximately 18 hours at 20 °C, with orbital shaking at 225 rpm (data not shown). Similar conditions were assumed to be suitable for the expression of the other CAR enzymes and indeed all CARs were well expressed. CARs were co-expressed with the Sfp phosphopantetheinyl transferase from *Bacillus subtilis* on a separate plasmid, as the loading of a phosphopantetheine group onto CAR enzymes has been shown to be essential for activity. [15]

### Kinetic characterization of CAR enzymes

The CAR enzymes were characterized in terms of their substrate specificity towards a range of aromatic carboxylic acids, a range of aliphatic unsaturated fatty acids, and the cofactors ATP and NADPH. Lists of the substrates with their chemical structure can be found in Figure 3 and Supplementary Figure 4. For each CAR, an initial assay at high substrate concentration (5 mM) was carried out to identify compounds for which CAR had activity. For those compounds where activity was detected, a full kinetic analysis was performed (Supplementary Figures 5-9). All the CARs that were tested showed similar $K_M$ values for NADPH and ATP. For NADPH the $K_M$ was between 24 and 36 μM, whilst for ATP $K_M$ values of between 64 and 84 μM were observed. These values are both well within the physiological ranges for these cofactors and in good agreement with previous studies. [6,10,23]. Production of benzaldehyde and 4-methylbenzaldehyde from the derivative

acids was confirmed by HPLC, with no other products observed. NADPH consumption was also confirmed as a good measure of aldehyde production (Supplementary figures 10 and 11).

**Investigating the effects of electronic density on aryl substituted carboxylic acid substrates**

All of the enzymes that we tested showed strong activity against the classical CAR substrate benzoic acid (compound **1**, Figure 3A, Table 2), which all previously studied CARs have shown activity against. [6,10,11] A series of substituents of varying electronic configuration were tested (compounds **2-5**, Figure 3A, Table 2). Compounds with more electron rich systems generally lowered $K_M$ giving increased catalytic efficiency compared to benzoic acid. Minimal activity was detected with 2- methoxybenzoic acid.

In contrast, compounds that incorporated an electron withdrawing nitro group in the benzene ring (compounds **6-8**, Figure 3B, Table 2) resulted in a large decrease in the turnover number of the CARs, in most cases inhibiting activity all together. Again, there was no detectable activity with a nitro group in the 2 position; whilst in the *para* position only tpCAR showed a low level of activity. However, all the CARs tested were active against 3-nitrobenzoic acid but with a lower $k_{cat}$ than benzoic acid. Absorbance at $OD_{340nm}$ by nitro compounds was shown not to interfere with the assay (Supplementary Figure 12).

**Investigating the effect of the aromatic unit on catalytic activity**

3-phenylpropionic acid (compound **9**, Figure 3C, Table 3) has a carboxylate group out of conjugation from the aryl ring, extended away from the aryl ring by two carbons giving the carboxylate group greater flexibility. When tested with the CARs, this change caused a reduction in $K_M$, with a similar or slightly lower $k_{cat}$. (*E*)-3-phenylprop-2-enoic acid (cinnamic acid; compound **10**, Figure 3C, Table 3), being a conjugated system, was expected to have activity between **9** and the model compound **1**, benzoic acid. The CAR activity against (*E*)-3-phenylprop-2-enoic acid showed a substantial reduction in $k_{cat}$ compared to 3-phenylpropionic acid or benzoic acid, with a slight further reduction in $K_M$.

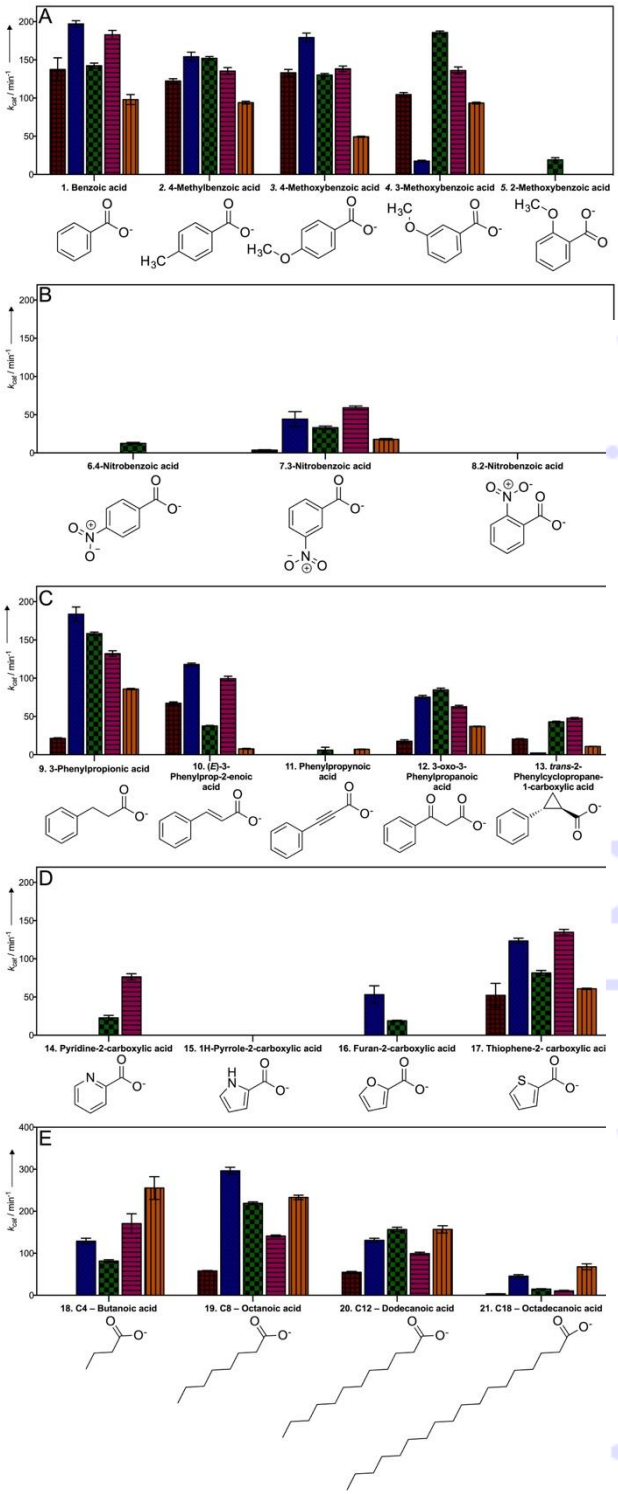


**Figure 3 – CAR activity for various benzoic acid derivatives, heterocycles and fatty acids.** The $k_{cat}$ ($min^{-1}$) determined for each enzyme against each substrate displayed as follows: mpCAR, msCAR, tpCAR, noCAR and niCAR. Below each substrate is its chemical structure. Error bars show the standard error. A: Benzoic acid and derivatives with electron donating groups. B: Derivatives with an electron withdrawing groups. C: Derivatives with various substituents between the carboxylate group and benzene ring. D: Heterocycles containing either an oxygen, sulfur or nitrogen. E: Fatty acids between four and eighteen carbons in length.

**Table 2** – CAR activity against benzoic acid and its derivatives with electron donating and withdrawing groups

| | | 1. Benzoic acid | 2. 4-Methylbenzoic acid | 3. 4-Methoxybenzoic acid | 4. 3-Methoxy benzoic acid | 5. 2-Methoxy benzoic acid | 6. 4-Nitrobenzoic acid | 7. 3-Nitrobenzoic acid | 8. 2-Nitrobenzoic acid |
|---|---|---|---|---|---|---|---|---|---|
| *Hammet sigma constants:* [28] | | *0* | *-0.17* | *-0.27* | *0.12* | - | *0.71* | *0.78* | - |
| **mpCAR** | $k_{cat}$ ($min^{-1}$) | 140 ± 20* | 122 ± 3 | 132 ± 4 | 104 ± 3 | NA | NA | 3.7 ± 0.5 | NA |
| | $K_M$ (mM) | 20 ± 4* | 3.7 ± 0.2 | 2.8 ± 0.2 | 3.0 ± 0.2 | NA | NA | 0.3 ± 0.1 | NA |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 7 ± 1* | 33 ± 2 | 48 ± 5 | 35 ± 3 | NA | NA | 11 ± 4 | NA |
| **msCAR** | $k_{cat}$ ($min^{-1}$) | 197 ± 4 | 154 ± 6 | 179 ± 6 | 18 ± 1* | NA | NA | 40 ± 10 | NA |
| | $K_M$ (mM) | 3.4 ± 0.2 | 0.16 ± 0.02 | 0.19 ± 0.02 | 12 ± 1* | NA | NA | 0.5 ± 0.2 | NA |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 57 ± 4 | 900 ± 100 | 930 ± 80 | 1.4 ± 0.2* | NA | NA | 100 ± 50 | NA |
| **tpCAR** | $k_{cat}$ ($min^{-1}$) | 142 ± 3 | 152 ± 2 | 130 ± 2 | 186 ± 2 | 19 ± 3 | 13 ± 1 | 33 ± 2 | NA |
| | $K_M$ (mM) | 2.0 ± 0.1 | 0.69 ± 0.03 | 0.45 ± 0.02 | 0.56 ± 0.02 | 9 ± 3 | 0.6 ± 0.1 | 0.7 ± 0.1 | NA |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 72 ± 6 | 220 ± 10 | 290 ± 10 | 334 ± 10 | 2.2 ± 0.7 | 22 ± 6 | 44 ± 8 | NA |
| **noCAR** | $k_{cat}$ ($min^{-1}$) | 183 ± 6 | 135 ± 5 | 138 ± 4 | 136 ± 5 | NA | NA | 59 ± 2 | NA |
| | $K_M$ (mM) | 2.1 ± 0.2 | 1.2 ± 0.2 | 1.1 ± 0.1 | 0.9 ± 0.1 | NA | NA | 2.5 ± 0.3 | NA |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 89.1 ± 8 | 110 ± 20 | 130 ± 10 | 150 ± 10 | NA | NA | 24 ± 3 | NA |
| **niCAR** | $k_{cat}$ ($min^{-1}$) | 98 ± 7 | 94 ± 2 | 49 ± 1 | 93 ± 1 | NA | NA | 18 ± 1 | NA |
| | $K_M$ (mM) | 0.9 ± 0.1 | 1.0 ± 0.1 | 0.25 ± 0.01 | 0.68 ± 0.03 | NA | NA | 5.6 ± 0.7 | NA |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 103 ± 9 | 97 ± 6 | 200 ± 10 | 137 ± 6 | NA | NA | 3.2 ± 0.4 | NA |

NA: no activity was detected with that substrate. *: $K_M$ was unusually large and substrates concentrations could not reach a high enough concentration to accurately determine kinetic constants. Errors represent the standard error. No Hammett constants are shown for 2-substituents as steric effects cannot be properly accounted for.

**Table 3** – CAR activity against benzoic acid derivatives with the carboxylic acid group extended from the ring.

| | | 9. 3-Phenylpropionic acid | 10. (*E*)-3-Phenylprop-2-enoic acid | 11. Phenylpropynoic acid | 12. 3-oxo-3-Phenylpropanoic acid | 13. *trans*-2-Phenylcyclopropane-1-carboxylic acid |
|---|---|---|---|---|---|---|
| **mpCAR** | $k_{cat}$ ($min^{-1}$) | 21.5 ± 0.7 | 67 ± 2 | NA | 18 ± 2 | 20 ± 1 |
| | $K_M$ (mM) | 3.0 ± 0.3 | 0.3 ± 0.02 | NA | 3.8 ± 0.8 | 1.8 ± 0.2 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 7.2 ± 0.7 | 240 ± 2 | NA | 5 ± 1 | 12 ± 1 |
| **msCAR** | $k_{cat}$ ($min^{-1}$) | 184 ± 9 | 118 ± 2 | NA | 75 ± 2 | 2.2 ± 0.1 |
| | $K_M$ (mM) | 0.16 ± 0.02 | 0.075 ± 0.004 | NA | 0.27 ± 0.02 | 0.006 ± 0.0001 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 1200 ± 200 | 1600 ± 500 | NA | 280 ± 20 | 380 ± 20 |
| **tpCAR** | $k_{cat}$ ($min^{-1}$) | 158 ± 2 | 38 ± 1 | 6 ± 4 | 85 ± 2 | 43 ± 1 |
| | $K_M$ (mM) | 0.32 ± 0.01 | 0.310 ± 0.002 | 0.09 ± 0.02 | 0.55 ± 0.04 | 0.061 ± 0.005 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 500 ± 20 | 120 ± 2 | 70 ± 40 | 150 ± 10 | 700 ± 60 |
| **noCAR** | $k_{cat}$ ($min^{-1}$) | 140 ± 4 | 105 ± 3 | NA | 63 ± 2 | 48 ± 1 |
| | $K_M$ (mM) | 2.7 ± 0.2 | 0.72 ± 0.07 | NA | 0.29 ± 0.03 | 1 ± 0.1 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 52 ± 4 | 147 ± 15 | NA | 210 ± 20 | 46 ± 3 |
| **niCAR** | $k_{cat}$ ($min^{-1}$) | 85.8 ± 0.9 | 7.7 ± 0.7 | 7 ± 0.4 | 37.1 ± 0.5 | 10.8 ± 0.3 |
| | $K_M$ (mM) | 0.97 ± 0.03 | 0.05 ± 0.02 | 1.3 ± 0.2 | 0.39 ± 0.02 | 0.21 ± 0.02 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 88 ± 3 | 170 ± 70 | 5 ± 1 | 94 ± 4 | 51 ± 5 |

NA: no activity was detected with that substrate. Errors represent the standard error.

**Table 4** – CAR activity against heterocycles and fatty acids

| | | 14. Pyridine-2-carboxylic acid | 15. 1*H*-Pyrrole-2-carboxylic acid | 16. Furan-2-carboxylic acid | 17. Thiophene-2-carboxylic acid | 18. C4 – Butanoic acid | 19. C8 – Octanoic acid | 20. C12 – Dodecanoic acid | 21. C18 – Octadecanoic acid |
|---|---|---|---|---|---|---|---|---|---|
| **mpCAR** | $k_{cat}$ ($min^{-1}$) | NA | NA | NA | 50 ± 20* | NA | 58 ± 1 | 55 ± 2 | 3.7 ± 0.3 |
| | $K_M$ (mM) | NA | NA | NA | 50 ± 20* | NA | 2.0 ± 0.1 | 0.09 ± 0.01 | 0.09 ± 0.02 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | NA | NA | NA | 1.1 ± 0.6* | NA | 29 ± 2 | 600 ± 70 | 39 ± 9 |
| **msCAR** | $k_{cat}$ ($min^{-1}$) | NA | NA | 50 ± 10 | 123 ± 4 | 129 ± 7* | 296 ± 8 | 131 ± 5 | 46 ± 4 |
| | $K_M$ (mM) | NA | NA | 13 ± 4 | 3.3 ± 0.3 | 7.9 ± 0.8* | 0.1 ± 0.01 | 0.05 ± 0.01 | 0.6 ± 0.09 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | NA | NA | 4 ± 2 | 37 ± 3 | 16 ± 2* | 3000 ± 300 | 2700 ± 400 | 80 ± 10 |
| **tpCAR** | $k_{cat}$ ($min^{-1}$) | 23 ± 3 | NA | 19 ± 1 | 82 ± 3 | 82 ± 3* | 219 ± 3 | 157 ± 5 | 15 ± 1 |
| | $K_M$ (mM) | 24 ± 7 | NA | 4.7 ± 0.5 | 3.3 ± 0.3 | 5.0 ± 0.4* | 0.2 ± 0.01 | 0.04 ± 0.01 | 0.12 ± 0.03 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 0.9 ± 0.3 | NA | 4.0 ± 0.4 | 25 ± 3 | 17 ± 2* | 1140 ± 50 | 3600 ± 400 | 120 ± 30 |
| **noCAR** | $k_{cat}$ ($min^{-1}$) | 76 ± 4 | NA | NA | 135 ± 4 | 170 ± 20* | 141 ± 2 | 99 ± 3 | 11 ± 1 |
| | $K_M$ (mM) | 20 ± 2 | NA | NA | 2.6 ± 0.2 | 50 ± 8* | 0.2 ± 0.01 | 0.04 ± 0.01 | 0.02 ± 0.01 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | 3.9 ± 0.4 | NA | NA | 52 ± 4 | 3.4 ± 0.7* | 750 ± 30 | 2500 ± 300 | 500 ± 300 |
| **niCAR** | $k_{cat}$ ($min^{-1}$) | NA | NA | NA | 60.8 ± 0.9 | 260 ± 30* | 233 ± 5 | 157 ± 9 | 68 ± 7 |
| | $K_M$ (mM) | NA | NA | NA | 1.00 ± 0.05 | 32 ± 4* | 0.2 ± 0.01 | 0.02 ± 0.01 | 0.7 ± 0.1 |
| | $k_{cat}$ / $K_M$ ($min^{-1}$ $mM^{-1}$) | NA | NA | NA | 58 ± 3 | 8 ± 1* | 1350 ± 90 | 7000 ± 2000 | 100 ± 20 |

NA: no activity was detected with that substrate. *: $K_M$ was unusually large and substrates concentrations could not reach a high enough concentration to accurately determine kinetic constants. Errors represent the standard error.

The cognate compound with a triple bond (phenylpropynoic acid; compound **11**, Figure 3C, Table 3) showed very low or no detectable activity in the CAR reaction.

Two other compounds were tested: firstly, the β-keto acid 3-oxo-3-phenylpropanoic acid (compound **12**, Figure 3C, Table 3) showed an increase in $K_M$ with mpCAR, msCAR and tpCAR, but a decrease in $K_M$ with noCAR and niCAR in comparison to 3-phenylpropionic acid. However, in all cases, the $k_{cat}$ was reduced when compared to 3-phenylpropionic acid or benzoic acid, as it was for (*E*)-3-phenylprop-2-enoic acid. Finally, *trans*-2-phenylcyclopropane-1-carboxylic acid (compound **13**, Figure 3C, Table 3) features a cyclopropane ring between the benzene ring and carboxylate group. For all the CARs this modification resulted in much lower $K_M$ values, and a much lower $k_{cat}$, compared to 3-phenylpropionic acid or benzoic acid.

### Heterocycles

Heterocycles containing nitrogen, oxygen or sulfur were tested (compounds **14-17**, Figure 3D, Table 4). Generally, weak activity was observed, with decreasing $K_M$ values for increasing heteroatom size. Where there was activity, $k_{cat}$ was generally lower than the activity observed with benzoic acid as substrate. In cases where no activity was detected it is possible that the $K_M$ was outside the range of detection of the assay.

### Fatty acids

All the CARs showed very high catalytic efficiency for fatty acids between eight and twelve carbons in length, with low $K_M$ values compared to benzoic acid (compound **18-21**, Figure 3E, Table 4). Octadecanoic acid (**21**), with a carbon chain length of 18 carbons, showed a similarly low $K_M$ but a greatly reduced $k_{cat}$. All CARs except mpCAR were active against butanoic acid (compound **18**, Figure 3E, Table 4) but with a very large $K_M$, in most cases too large to characterize accurately. In general, mpCAR was much less efficient with fatty acids than the other CAR enzymes.

### Effects of pH

The activity of an enzyme at different pH values is an important consideration for an industrial enzyme. Therefore, the effect of pH on CAR activity was examined by measuring activity against

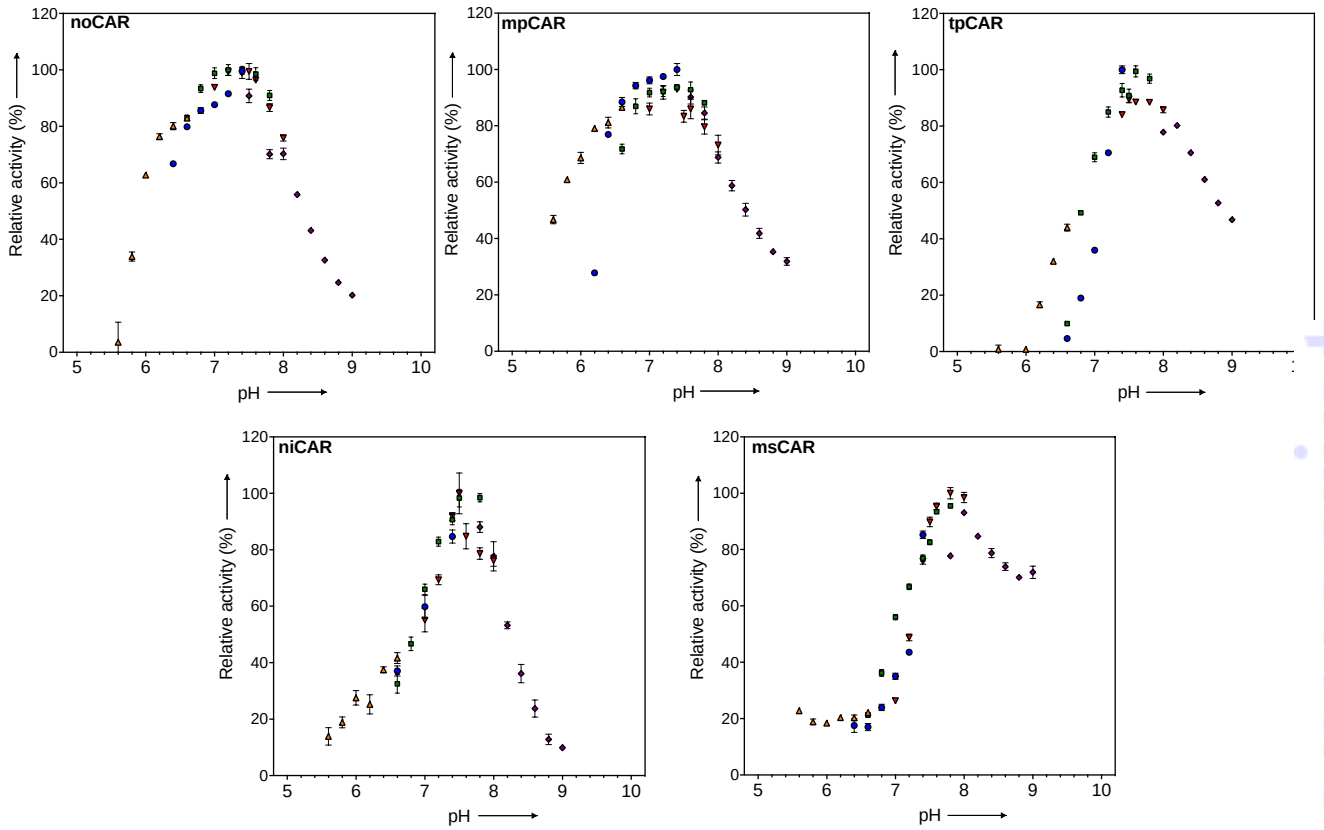


**Figure 4 – The activity of CAR enzymes in response to pH.** Overlapping buffers were used to cover a range from pH 5.6 to pH 9.0 in intervals of 0.2 and are displayed as follows: MES-NaOH ▲ PIPES-NaOH ●, MOPS-NaOH ■, HEPES-NaOH ▼, Tris-HCl ◆ Activity against 4-methylbenzoic acid is shown relative to the highest activity at 100%. Errors bars show the combined standard deviation of three readings and three blank readings (with no enzyme) at each pH

benzoic acid at different pH values. mpCAR, niCAR, noCAR and tpCAR all showed optimum activity at pH 7.5, whilst msCAR showed an optimum activity at pH 7.8 (Figure 4). Both niCAR and tpCAR show a sharp peak of activity around pH 7.5, with activity quickly decreasing as the pH moved away from this point. In contrast, mpCAR and noCAR show a slightly broader optimum around pH 7.0 to 7.6. msCAR behaves very differently from the other CARs. At more acidic pH values between pH 5.5 and 6.8 it shows very low activity where the other CARs are more active. However, it is also more active at more alkaline pH values where the other CARs are less active.

### Effects of temperature

Thermostability was investigated by incubating the CAR enzymes at various temperatures for half an hour and measuring residual activity against 4-methylbenzoic acid relative to a control kept on ice. tpCAR was the least thermostable CAR tested, being completely inactive after half an hour at 42 °C (Figure 5A). In contrast, mpCAR, a CAR from the moderate thermophile *M. phlei,* retains 92% of its activity following the same incubation at 42 °C. mpCAR was able to retain residual activity up to 50 °C making it the most thermostable CAR identified to date. Both niCAR and noCAR showed intermediate thermostability, denaturing at temperatures beyond 44 °C, while msCAR is marginally more thermostable and is able to retain some activity until 47 °C.

Activity at temperature was tested in a 10 minute reaction. The more thermostable CARs, mpCAR, msCAR and niCAR all showed an optimum activity of 42 °C (Figure 5B). Activity decreased past this temperature at various degrees relative to the thermostability of each enzyme. noCAR showed a slightly lower optimum at 38 °C while tpCAR had a much lower optimum still, at only 31 °C.

The half-life and degradation constant at 30 °C were calculated by measuring activity at various time points over 120 hours. The data were fitted to a one phase decay equation by non-linear least squares regression. mpCAR, a CAR from a moderate thermophile, showed by far the longest half-life at 30 °C at 123.2 hours (Table 5). In contrast, tpCAR has a much shorter half-life of only 25.0 hours. The half-lives of msCAR, niCAR and noCAR fell between these extremes at 53.7, 42.9 and 35.3 hours respectively. Total turnover numbers (TTN) for the three best substrates were calculated as $k_{cat}$ / $K_D$ (Table 5).

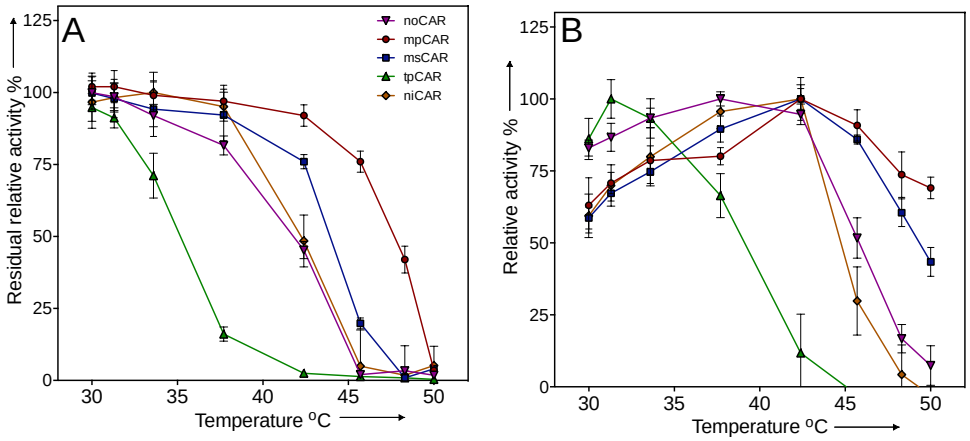


**Figure 5 – The effects of temperature on CAR enzymes**. A – Thermostability of CAR enzymes. The residual activity of CAR enzymes against 4-methylbenz acid after a 30-minute incubation at different temperatures is displayed. Activity is shown relative to a control sample kept at 4 °C with errors bars showing the standard deviation of three readings. B – Activity of CAR enzymes at different temperatures. Activity is relative to the fastest rate at 100%. Error bars show the combined standard deviation of three readings and three blank reading (with no substrate) at each temperature.

### Product inhibition

mpCAR was tested for product inhibition with AMP, $NADP^+$ and $PP_i$. $NADP^+$ showed competitive inhibition with NADPH with a $K_I$ of 143 ± 8 μM (Supplementary Figure 13) and AMP was a competitive inhibitor of ATP with a $K_I$ of 8200 ± 900 μM (Supplementary Figure 14). $PP_i$ showed mixed inhibition with ATP with a $K_I$ of 220 ± 50 μM, and an α of 2.5 ± 1.4. Surprisingly, $PP_i$ also showed competitive inhibition with 4-methylbenzoic acid with a $K_I$ of 340 ± 40 μM (Supplementary Figures 15 and 16).

**Table 5 – Half-life and degradation constant $K_D$ of CAR enzymes when incubated at 30 °C.**

| Enzyme | Halflife (hours) | $K_D$ (hrs $^{-1}$) | TTN Benzoic acid | TTN 4-Methylbenzoic acid | TTN 4-Methoxybenzoic acid |
|---|---|---|---|---|---|
| mpCAR | 123.2 | 0.0056 ± 0.004 | 30000 ± 20000 | 20000 ± 20000 | 20000 ± 20000 |
| msCAR | 53.7 | 0.013 ± 0.001 | 15000 ± 1000 | 10000 ± 1000 | 14000 ± 1000 |
| niCAR | 25 | 0.28 ± 0.002 | 350 ± 30 | 336 ± 8 | 175 ± 4 |
| noCAR | 35.3 | 0.02 ± 0.002 | 9000 ± 1000 | 6800 ± 700 | 6900 ± 700 |
| tpCAR | 42.9 | 0.016 ± 0.002 | 9000 ± 1000 | 6000 ± 1000 | 8000 ± 1000 |

The halflife, $K_D$ and TTN for the three best substrates of CAR enzymes calculated from activity after incubation at 30 °C over time, fitted to $Y=Y0*e^{-K*X}$. Standard error for $K_D$ is shown. TTN has been calculated as *kcat* / $K_D$, with the combined error shown.

## Discussion

The CAR enzymes offer an excellent opportunity for green chemistry: they offer the opportunity to reduce carboxylic acids selectively to aldehydes, without the use of harsh reducing agents CARs also have the clear advantage over other enzymes capable of carrying out this reaction of the reduced product being thermodynamically favored, due to the hydrolysis of ATP. Although previous studies have identified a few CARs from different species, and demonstrated that they have activity against diverse acids, none of these studies has provided a detailed, kinetic comparison of diverse CARs. We therefore aimed to thoroughly characterize example CARs from across the known CAR family, together with the best-characterized CAR from *N. iowensis*. Our aim was to demonstrate the similarities and differences between these CARs, learn more about the CAR mechanism, and highlight the potential of these enzymes for biocatalysis.

### The effect of the addition of electron donating or withdrawing groups

The reduction of carboxylic acids to aldehydes typically involves a transfer of a 'hydride' to the carbonyl unit. We therefore initially expected that electron withdrawing groups, which make this carbon more electrophilic, would be preferred substrates. However, our observation was that contrary to our expectation, electron donating groups were preferred substrates (Figure 3, Table 2). The addition of electron donating groups to benzoic acid resulted in a reduction in $K_M$, and so an increase in catalytic efficiency. We reasoned that these groups would drive electrons into the π-system, making the first step of the reaction (attack by the negatively charged carboxylate group on the α phosphate of ATP) more favorable. As two of the other steps (2 and 4) involve nucleophilic attacks on the acid

group carbon atom of the carboxyl group (which should favor electron withdrawing groups), this strongly suggests that the first step in the reaction has the greatest impact on substrate specificity and selection. It is possible that the reduced $K_M$ with electron donating substituents is a consequence of the acyl-AMP intermediate forming more readily, although very detailed studies of the kinetics of this individual step would be required to confirm this. Indeed, previous studies of NRPSs have shown the adenylation reaction to be a rate limiting step.[33] In long-chain fatty acid ligases, the acyl-AMP intermediate has been shown to be unable to leave the active site [19], so the addition of a group which likely improves the formation of this intermediate might be expected to cause a lower $K_M$ and greater catalytic efficiency. Moreover, when benzyl-AMP was used as a substrate with a CAR from *Nocardia asteroides* it showed a $K_M$ of 70 nM, compared to 260 nM for benzoic acid, suggesting that this intermediate binds more tightly to the enzyme. [22] Furthermore, the phosphopantetheine binding and C-terminal reductase domains shows high sequence identity to that of other ANL superfamily members that process very different substrates. For example a NRPS from *Mycobacterium intracellulare*, WP_014382786.1, has an average of 58 % identity to the CARs in Figure 2 for this C-terminal region. This strongly suggests that substrate specificity must be determined in the adenylation domain, likely at the formation of the first intermediate.

In the 3-position (**4**), the methoxy group has no resonance effect on the carboxylic acid and so is actually slightly electron withdrawing by induction, as indicated by the Hammett sigma constants in Table 2. In many of the CARs, the $k_{cat}$ of 3-substituted benzoic acids shows a small reduction compared to 4-substituted acids, with msCAR showing greatly reduced activity. However, these are still good substrate for most of the CARs. It is likely that there are further interactions between the substrate and the active site binding pocket, and that electronic effects alone cannot account for all differences in activity.

Very low or no activity was found with the 2-methoxy substituent of benzoic acid (**5**). This suggests that there is a steric interference by the methoxy group on the binding of the nearby carboxylate group to the relevant area of the active site. This effect has been reported for other CARs examined to date with other 2-substituents. However, some cases suggest there is activity, but at a low level.[22,23] No structure of a CAR enzyme has yet been described, and this would be highly beneficial in understanding the effects of groups in the 2-position.

All substrates with an electron withdrawing group (**6-8**) showed much lower $k_{cat}$ values than benzoic acid, in most cases inhibiting activity all together. These groups should increase the propensity of the carbonyl carbon to nucleophilic attack in steps 2 and 4 of the reaction. Therefore, this strongly suggests again that these two steps are of limited relevance for substrate specificity. Only the 3-nitro substituent (**7)** showed activity with all the CAR enzymes, likely as in this position the electron withdrawing group has no resonance effect on the carboxylate group. As is the case with the methoxy group, it is possible that a 2-nitro substituent (**8**) inhibits activity due to a steric hindrance because of its close proximity to the carboxylate group.

Previously it has been reported that 2-substituted benzoic acids are poor substrates for niCAR, in good agreement with our data.[23] However very low activity was observed with **8** previously, which we did not detect. Substrates with the addition of electron donating groups to benzoic acid were previously shown to be good substrates for niCAR, in agreement with our results. The activity of niCAR with electron withdrawing chloro and bromo 3-substituted benzoic acids supports our reasoning that in the 3-position the absence of a resonance effect allows better activity with these substrates than in the other positions. [23]

**Modification of the aromatic ring and unit on CAR activity.**

The CARs generally showed less activity towards heterocycles compared to a benzene ring. They showed a preference for heterocycles containing a larger heteroatom, or with a less aromatic nature. In substrate **17**, the lone pairs of electrons in the sulfur atom are more dispersed and less available for bonding, which possibly results in the lower $K_M$. In contrast, the nitrogen atom in substrates **14** or **15** has lone pairs more available for bonding, which may result in the very large $K_M$ values, or lack of activity observed. Substrate **16**, with an oxygen atom in the heterocycle, sits between these substrates in both respects.

Extension of the carboxylate group away from the aryl group in **9** disrupts its influence on the carboxylic acid, and makes a less sterically rigid substrate. This difference seems to have made the carboxylic acid group more accessible, as the $K_M$ is much lower than substrate **1** in most cases. In contrast, the inclusion of a double bond in substrate **10**, should withdraw electrons from the carboxylic acid group. This would be beneficial for nucleophilic attack on the carbonyl in steps 2 and 4 of the reaction, but detrimental to the initial attack by the oxygen of the carboxylate group on ATP. The result is a significant drop in $k_{cat}$ compared to substrate **1**. The double bond also makes the molecule more rigid in an apparently favorable conformation, as the $K_M$ is even smaller than for substrate **9**. When a triple bond is added to the substrate (**11**), the molecule is very rigid and flat, with a more electron deficient carboxylic acid group. These effects together removed activity in nearly all the CARs. The presence of a β-ketone group into the β-carbon of substrate **9** will have a similar effect to the inclusion of a double bond in substrate **10**, very weakly withdrawing electrons from the carboxylic acid group. The ketone group (substrate **12**) had mixed effects on the $K_M$ for the various CARs, suggesting differing interactions taking place with the ketone group within the active sites of the enzymes. These observations agree with the hypothesis that the first step of the proposed reaction mechanism is rate limiting.

### Fatty acids

Fatty acids make interesting substrates, since fatty alcohols can be used as biofuels, in detergents, surfactants and polymers [7]. As was observed for the CAR from *Mycobacterium marinum* (mmCAR), most of the CARs tested were active against fatty acids between C4 and C18, with similar kinetics to previous work observed [10]. Catalytic efficiency with substrate **18** (C4) was very poor, primarily due to large $K_M$'s for this substrate suggesting it might be too small to make the necessary interactions in the active site of the adenylation domain. However, larger fatty acids showed much lower $K_M$'s, with high turnover numbers resulting in catalytic efficiencies higher than any of the aromatic substrates tested in many cases. As the acyl chain length increased past substrate **19** (C8) $k_{cat}$ decreased, reaching a low residual level for substrate **21** (C18). Both niCAR and msCAR showed a better turnover number with substrate **21** than the other CARs, suggesting these enzymes might be better suited to larger substrates. Recently two other CARs, in combination with niCAR and mmCAR have been shown to have activity against ethanoic, butanoic, 2-methyl butanoic and 2-oxobutanoic acids, highlighting that CARs can accept small fatty acids and that they can tolerate the addition of groups such as a methyl or carbonyl group onto the alpha carbon. However, 2-aminobutanoic acid was also tested but showed no activity. [11]

### Effects of pH and temperature

The operating pH and temperature range of an enzyme is an important consideration for a potential biocatalyst. Stability at extremes of pH and in solvents are characteristics often found in thermostable proteins, as the mechanisms stabilizing these proteins against high temperature can also be stabilizing against these other conditions. We observed an optimum pH of 7.5 for four of the five CARs tested, with a general tolerance to acidic pH, consistent with previously reported data on the activity of other CARs.[22] In particular, both mpCAR and noCAR were able to tolerate pH 6 with only a small loss of activity (whilst other CARs showed a much narrower optimum). In contrast to this, msCAR is clearly better suited to more alkaline pH values (Figure 4). This therefore offers a CAR suitable for use in biocatalysis in conjunction with other enzymes favoring a similarly alkaline pH.

mpCAR showed by far the best thermostability of any characterized CAR (Figure 5, Table 5). We also observed that it shows a much lower catalytic efficiency in general than the other CARs at 30 °C (Tables 2-4). Possibly, there has been a trade-off between the rigidity of the enzyme (providing thermostability) and flexibility to allow a broader substrate range. It was notable that the rate enhancement in mpCAR at its optimum temperature compared to 30 °C was little greater than that for other CARs (Figure 5B). In contrast, tpCAR shows very poor thermostability (Figure 5), but is active with many of the substrates that the others CARs could not turn over (e.g. compounds **5**, **6**, **11**, **14**). A possible compromise enzyme is msCAR, which shows the next best thermostability, and also has generally good catalytic efficiency. When choosing an enzyme for industrial use, the lifespan of the enzyme can be an important consideration. TTN can be calculated as a measure of how effective an enzyme will be over its lifetime, which we have demonstrated with three of the best CAR substrates (Table 5). In this respect the most thermostable CARs have an obvious advantage in that the total turnover number of these enzymes will be much greater.[29] We observed that the lifespan of the enzyme at 30 °C (Table 5) mirrored the thermostability of the enzymes exactly (Figure 5), suggesting that a test of thermostability will be a good predictor of lifespan for CARs.

The CAR enzymes in this study show only moderate thermostability. To date no CAR enzymes have been identified in any thermophilic organisms. A thermostable CAR enzyme would be attractive for use industrially as this enzyme would likely be resistant to other denaturing forces such as extremes of pH or organic solvent and likely offer a higher total turnover number for use *in vitro* reactions.

### Product inhibition and reaction mechanism

mpCAR was shown to be inhibited by most of its reaction products and it is assumed that the other CARs share this inhibition. It is unsurprising that $NADP^+$ acts as a competitive inhibitor of NADPH (Supplementary Figure 13) as $NADP^+$ is likely also able to bind to the Rossmann fold of the reductase domain. AMP acts as a competitive inhibitor against ATP (Supplementary Figure 14), likely as they are very similar molecules. AMP has also been shown to be a competitive inhibitor of ATP in long-chain fatty acid CoA synthetases, in which the adenylation domain shows significant homology to the CAR adenylation domain.[30]

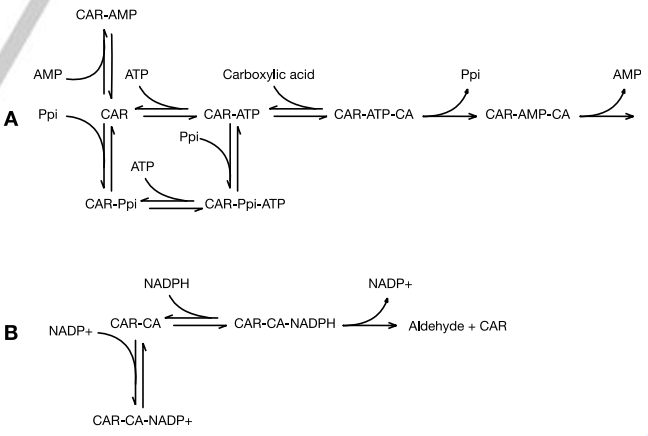


**Figure 6 – Model for binding of substrates and inhibitors to the CAR enzyme. A**: binding and release of substrates, products and inhibitors in the adenlyation domain. The final result is the formation of a thioester intermediate with the phosphopantetheine arm, represented by CAR-CA. The phosphopantetheine arm can then transfer CA to the reduction domain, **B**, where it is reduced by NADPH, releasing the aldehyde product

$PP_i$ showed mixed inhibition against ATP but competitive inhibition against 4-methylbenzoic acid (Supplementary Figures 15 and 16). This pattern of inhibition is characteristic for ordered sequential bisubstrate reactions.[31] This indicates that ATP is first to bind to the adenylation domain and is then followed by a carboxylic acid. Long-chain fatty acid CoA synthetases show the same ordered binding of these substrates.[19] We therefore propose a model for the ordered binding of substrates and inhibitors to the CAR enzyme based on these results (Figure 6). It is also interesting that whilst $PP_i$ is a product of ATP, its activity as an inhibitor shows that it preferentially binds to the carboxylic acid binding site. CARs might therefore need to be combined with other enzymes such as phosphite dehydrogenase[32] or inorganic pyrophosphatase[33] to overcome product inhibition in an industrial process. Indeed the *in vitro* turnover of niCAR has been shown to be improved by the addition of an inorganic pyrophosphatase enzyme.[34]

### CAR phylogeny and insight into the adenylation step

Here we have provided the first glimpse of CAR evolution within the *Actinomycetes*. From the phylogeny it can be hypothesized that the CARs may have propagated through the *Nocardia* and *Mycobacteria* by a series of early horizontal transfer events. This is most apparent in *M. smegmatis*, which possesses three CAR paralogues that cluster in two distinct Mycobacterial clades. Additionally, it is apparent that a large amount of change has occurred within the *Tsukamurella.* This could reflect the slightly more promiscuous substrate range of tpCAR.

We presented evidence that the adenylation domain of the CARs belongs to the ANL superfamily of enzymes due to the presence of conserved hallmark motifs. Both nonribosomal peptide synthetases (NRPSs) and the acyl-CoA synthetases similarly use an acyl group to form a thioester between a substrate and a pantetheine thiol, supporting this interpretation. Furthermore, the NPRSs mobilize their substrate following thiolation of a phosphopantetheine arm bound to a *holo*-acyl carrier protein domain. Parallels can be drawn between both above reactions and the proposed mechanism of CAR activity in Figure 1. This offers the opportunity to exploit the extensive studies on the ANL superfamily to gain insight into the finer details of the mechanism of carboxylic acid reduction employed by CARs.

In particular, ANL superfamily members are further partitioned into two subdomains – a large (~450 aa) N-terminal domain and a small (~100 aa) C-terminal domain, connected by a flexible linker. Crystal structures show the substrate-binding pocket is formed by the N- and C-domain interface. Substrate adenylation proceeds in a two-step manner, where following the formation of an acyl-bound intermediate and the release of $PP_i$, the active site undergoes large conformational changes due to a ~140° rotation of the C-terminal domain. Within the NRPSs and the acyl-CoA synthetases, the second domain architecture facilitates thiolation of the phosphopantetheine. Lysines that are required within each active site are positioned on opposing faces of the C-terminal domain, and are conserved within the CARs (Supplementary Figure 1).[35] This suggests that the CARs also undergo characteristic ANL superfamily domain-alteration between steps 1 and 2 (figure 1) to catalytically isolate the adenylation and thioester forming reactions.[18]

### Summary

CARs have been proposed as a useful tool for novel biocatalysis. This study has demonstrated that, across the entire extant phylogeny of CARs, similar substrates are preferred by this family of enzymes, with some enzymes being more promiscuous than others. In particular, our detailed kinetic analysis of CARs strongly suggests that the first step in the proposed reaction mechanism, during which an AMP-carboxylic acid phosphoester intermediate is formed with the release of $PP_i$, is critical for determining suitable substrates. Consequently, the addition of groups that donate electrons, making the oxygen of the carboxylic acid more electronegative, will be better substrates; and aliphatic acids are strongly preferred to aromatic acids. This study also highlighted that, similarly to other members of the ANL superfamily to which the CARs belong, this first step is an ordered sequential Bi Bi reaction, with ATP being bound before the carboxylic acid. Of particular relevance to biocatalysis is that all of the by-products of the reaction ($PP_i$, AMP and $NADP^+$) appear to be inhibitors: for the use of CARs *in vitro* there is a need to remove or regenerate these. These data further validate CARs as a useful tool for novel biocatalytic reactions, and highlight their potential when integrated with other enzymes *in vitro* for efficient reduction of carboxylic acids to aldehydes.

## Methods

#### Alignments and Phylogeny Construction

Unless specified, all algorithms were performed under default settings. 48 sequences were retrieved by homology search in BLAST to the *N. iowensis* CAR. Alignments were performed using the MUSCLE plug-in within Geneious version 9.1 (http://www.geneious.com).[36] Sequence masking was conducted with the Gblocks algorithm within the Phylogeny.fr online tool (http://www.phylogeny.fr).[37] ProtTest (version 3.4)[38] analysis of the aligned dataset was performed in the command line. MrBayes (version 3.2.6)[39] was run in the command line as follows: The amino acid substitution model was fixed to WAG with a gamma-distributed rate variation across a proportion of invariable sites and 8 gamma categories. The analysis was run for 1,000,000 MCMCMC generations, sampling every 100 generations with two parallel runs and four chains (containing one heated chain of temperature 0.2), with a burn-in of 25%. Trees were visualised, midpoint rooted and modified in FigTree version 1.4 (http://tree.bio.ed.ac.uk/software/figtree/).

It must be noted that a more complete list of 124 CAR homologues was retrieved (Supplementary Figures 17 and 18). However, a reduced set of sequences was used as this allowed the construction of a more reliable phylogeny.

**Expression and purification**

CAR genes (except niCAR) were cloned into expression vectors pNIC28-Bsa4 [28] or obtained from Prozomix, cloned into pET28a (Novagen). A pET plasmid for the expression of niCAR was obtained from Andrew Hill (University of Manchester). All contained a N-terminal 6x histidine tag.[28] Full sequences for all vectors are supplied as Supplementary information. Vectors were transformed into BL21 (DE3) *E. coli* along with a pCDF-Duet1 vector containing a phosphopantetheine transferase from *Bacillus subtilis* for its co-expression with the CARs. Expression was carried out in LB media with the addition of 50 μg/μl each kanamycin and spectinomycin. Cells were grown to approximately 0.6 $OD_{600nm}$ at 37 °C with shaking at 225 rpm, at which point IPTG was added to a concentration of 150 μM and temperature was dropped to 20 °C for protein expression overnight. Cells were harvested by centrifuging 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 to remove the insoluble fraction.

CARs 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 approximately 2 mg/ml. To calculate protein concentration from $OD_{280nm}$, an extinction coefficient and molecular weight for each enzyme was calculated using the ExPaSy ProtParam tool, and are shown in Supplementary Figure 19. Yields of approximately 2 – 10 mg purified protein per liter of culture were obtained, with 2 – 4 L of culture prepared per batch. Single use aliquots of protein were stored at -80 °C.

**Standard enzyme assay**

Unless otherwise specified, assays were carried out in 100 mM Tris-HCl pH 7.5 prepared at 30 °C, 1 mM ATP, 0.25 mM NADPH, 10 mM $MgCl_2$, 2-6 μg of purified CAR enzyme and 5 mM carboxylic acid substrate in a total volume of 200 μl. Carboxylic acid substrates were prepared in DMSO at 500 mM. The oxidation of NADPH was used to monitor the reactions by measuring the absorbance of NADPH at 340 nm. Reactions were performed in triplicate in a 96-well microtitre plate using a Tecan M200 plate reader at 30 °C over the course of 5 or 10 minutes, after a 5 minute preincubation at 30 °C. Where convenient, an EpMotion 7050 (Eppendorf) liquid handling robot was used to set up the assays.

**Kinetic analysis of substrate specificity.**

Kinetic analysis was performed by picking eight appropriate substrate concentrations around an approximate $K_M$ value for each substrate, and measuring initial rates as described previously. Rates were fitted to the Michaelis-Menten equation by non-linear least squares regression using GraphPad Prism 5.0. To calculate constants for ATP and NADPH, 5 mM (*E*)-3-phenylprop-2-enoic acid was used as the carboxylic acid substrate, except for niCAR where 5 mM 4-methylbenzoic acid was used.

**pH vs activity.**

Buffers were prepared and titrated to the correct pH using NaOH or HCl whilst at 30 °C, covering pH values in intervals of 0.2. The buffers 50 mM MES pH 5.6 to 6.6, 50 mM PIPES pH 6.4 to 7.4, 50 mM MOPS pH 6.6 to 7.8, 50 mM HEPES pH 7.0 to pH 8.0 and 50 mM Tris pH 7.8 to pH 9.0 were used. Reactions were carried out as standard with 1 mM ATP, 0.25 mM NADPH, 10 mM $MgCl_2$ 2-6 μg of purified CAR enzyme and 5 mM 4-methylbenzoic acid. Blanks containing no enzyme were used to subtract a blank rate at each pH value. Initial rates were calculated as relative activity against the fastest result at 100 %.

**Thermostability**

A solution containing 2 μg of purified enzyme, 0.25 mM NADPH, 1 mM ATP, 10 mM $MgCl_2$, 100 mM Tris-HCl pH 7.5 was incubated across the temperature gradient of a Biorad thermocycler from 30 °C to 50 °C for 30 minutes. The sample was cooled and assayed for CAR activity against 4-methylbenzoic acid, in comparison to a control sample that remained on ice.

**Degradation at 30 °C**

2 ml samples at 2 mg/ml in 25 mM HEPES, pH 7.5, 0.1 M NaCl were incubated at 30 °C over a 120 hour period. At specified time intervals, samples were taken and assayed for enzyme activity against 4-methylbenzoic acid. Rates were calculated relative to the first reading at 100 %, and fitted to a model of first order thermal deactivation using the equation $Y = Y0 * e^{-K*X}$ where Y is the relative activity and X is the time in hours.

**Temperature vs activity**

100 mM Tris-HCl pH 7.5 was prepared at assay temperatures between 30 and 50 °C. Assays were performed as for the thermostability experiment using the temperature gradient of a Biorad thermocycler from 30 °C to 50 °C over the course of 10 minutes, before rapidly cooling on ice with the addition of 10 mM NaOH. A blank reaction with no substrate was used to calculate the NADPH used in the reaction. Activity was calculated relative to the maximum rate at 100 %.

**Product inhibition**

Potential inhibitors were titrated across a broad range of concentrations to determine whether inhibition occurred and to give an idea of an approximate $K_I$. Kinetic analysis then was performed as described above using substrates that each inhibitor was likely competitive against, with the addition of the inhibitors at a range of concentrations based around the approximate $K_I$. Data were fitted using GraphPad Prism 5.0 by non-linear least squares regression to different models of enzyme inhibition. The model with the best fit for the data was used to determine the mode of inhibition. Where inhibition was not competitive, additional analysis was carried out with other substrates.

## Acknowledgements

The authors thank Andrew Hill (University of Manchester) for providing many of the substrates tested, the pCDF-Sfp plasmid and the plasmid for the expression of niCAR; and Clive Mountain (GSK), Stacy Clark (GSK), and Alison Hill (University of Exeter) for advice on the chemistry of the CAR reaction, and Jennifer

Farrar (Georgia Institute of Technology) for providing walltime on her server to run the Bayesian analyses. Nzomics (Prof. Gary Black and team) and Prozomix (Simon Charnock and team) are gratefully acknowledged for cloning msCAR and tpCAR. WF was funded by BBSRC (grant no. BB/K501001/1) and GlaxoSmithKline; AT was funded by BBSRC (grant no. BB/J014400/1). Requests for raw data should be sent to NJH.

[1] Y.-H. P. Zhang, *Biotechnol. Adv.* **2015**, *33*, 1467–1483.

[2] J. A. Littlechild, *Biochem. Soc. Trans.* **2011**, *39*, 155–8.

[3] H.-P. Meyer, E. Eichhorn, S. Hanlon, S. Lütz, M. Schürmann, R. Wohlgemuth, R. Coppolecchia, *Catal. Sci. Technol.* **2013**, *3*, 29–40.

[4] A. S. Kertes, C. J. King, H. W. Blanch, *Biotechnol. Bioeng.* **2009**, *103*, 431–445.

[5] B. Erickson, Nelson, P. Winters, *Biotechnol. J.* **2012**, *7*, 176–85.

[6] A. He, T. Li, L. Daniels, I. Fotheringham, J. P. N. Rosazza, *Appl. Environ. Microbiol.* **2004**, *70*, 1874–1881.

[7] K. Napora-Wijata, G. A. Strohmeier, M. Winkler, *Biotechnol. J.* **2014**, *9*, 822–843.

[8] J. Heider, K. Ma, M. W. W. Adams, *J. Bacteriol.* **1995**, *177*, 4757–4764.

[9] F. Hollmann, Y. Ni, P. L. Hagedoorn, I. W. C. E. Arends, J.-H. Xu, *Chem. Commun.* **2012**, 12056–12059.

[10] M. K. Akhtar, N. J. Turner, P. R. Jones, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 87–92.

[11] M. Moura, D. Pertusi, S. Lenzini, N. Bhan, L. J. Broadbelt, K. E. J. Tyo, *Biotechnol. Bioeng.* **2015**, *113*, 944–952.

[12] 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.

[13] P. Kallio, A. Pasztor, K. Thiel, M. K. Akhtar, P. R. Jones, *Nat. Commun.* **2014**, *5*, 4731.

[14] A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, et al., *Nucleic Acids Res.* **2015**, *43*, D222-6.

[15] P. Venkitasubramanian, L. Daniels, J. P. N. Rosazza, *J. Biol. Chem.* **2007**, *282*, 478–485.

[16] D. E. Ehmann, A. M. Gehring, C. T. Walsh, *Biochemistry* **1999**, *38*, 6171–6177.

[17] D. I. I. Chan, H. J. J. Vogel, *Biochem. J.* **2010**, *430*, 1–19.

[18] A. M. Gulick, *ACS Chem. Biol.* **2009**, *4*, 811–827.

[19] Y. Hisanaga, H. Ago, K. Hamada, K. Ida, M. Yamamoto, T. Hori, M. Sugahara, S. Kuramitsu, S. Yokoyama, M. Miyano, *J. Biol. Chem.* **2004**, *279*, 31717–26.

[20] G. G. Gross, M. H. Zenk, *Eur. J. Biochem.* **1969**, *8*, 413–419.

[21] D. Schwendenwein, G. Fiume, H. Weber, F. Rudroff, M. Winkler, *Adv. Synth. Catal.* **2016**, *In Press*, DOI 10.1002/adsc.201600914.

[22] N. Kato, E.-H. Joung, H.-C. Yang, M. Masuda, M. Shimao, H. Yanase, *Agric. Biol. Chem.* **1991**, *55*, 757–762.

[23] T. Li, J. P. N. Rosazza, *J. Bacteriol.* **1997**, *179*, 3482–3487.

[24] E. Brenna, F. Cannavale, M. Crotti, F. Parmeggiani, A. Romagnolo, F. Spina, G. C. Varese, *J. Mol. Catal. B Enzym.* **2015**, *116*, 83–88.

[25] M. Winkler, C. K. Winkler, *Monatshefte fur Chemie* **2016**, *147*, 575–578.

[26] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651–2674.

[27] N. Lodders, P. Kämpfer, *eLS* **2007**, DOI 10.1002/9780470015902.a0020392.

[28] C. Hansch, S. D. Rockwell, P. Y. C. Jow, A. Leo, E. E. Steller, *J. Med. Chem.* **1977**, *20*, 304–306.

[29] T. A. Rogers, A. S. Bommarius, *Chem. Eng. Sci.* **2011**, *65*, 2118–2124.

[30] H. Li, E. M. Melton, S. Quackenbush, C. C. DiRusso, P. N. Black, *Biochim. Biophys. Acta* **2007**, *1771*, 1246–1253.

[31] P. F. Cook, W. W. Cleland, *Enzyme Kinetics and Mechanism*, Garland Science, New York, **2007**.

[32] Y. Zou, H. Zhang, J. S. Brunzelle, T. W. Johannes, R. Woodyer, J. E. Hung, N. Nair, W. a. Van Der Donk, H. Zhao, S. K. Nair, *Biochemistry* **2012**, *51*, 4263–4270.

[33] R. Lahti, *Microbiol. Rev.* **1983**, *47*, 169–178.

[34] A. M. Kunjapur, B. Cervantes, K. L. J. Prather, *Biochem. Eng. J.* **2016**, *109*, 19–27.

[35] J. A. Sundlov, D. M. Fontaine, T. L. Southworth, B. R. Branchini, A. M. Gulick, *Biochemistry* **2012**, *51*, 6493–6495.

[36] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, et al., *Bioinformatics* **2012**, *28*, 1647–1649.

[37] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, et al., *Nucleic Acids Res.* **2008**, *36*, 465–469.

[38] F. Abascal, R. Zardoya, D. Posada, *Bioinformatics* **2005**, *21*, 2104–2105.

[39] F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, *Syst. Biol.* **2012**, *61*, 539–542.

**Table of Contents:**

Carboxylic acid reductases (CARs) are enzymes that reduce such acids to aldehydes.

Example CARs from across the phylogeny have been characterized.

CARs generally prefer electron rich substrates.

Inhibition studies support an ordered binding of substrates for the first catalytic step.

**Keywords:** carboxylic acid reductase • biocatalysis • enzyme • green chemistry • oxidoreductase

10.1002/cctc.201601249