# Rational approaches for engineering novel functionalities in carbon-carbon bond forming enzymes

Perrin Baker [a], Stephen Y. K. Seah [a,*]

**Abstract:** Enzymes that catalyze carbon-carbon bond formation can be exploited as biocatalyst for synthetic organic chemistry. However, natural enzymes frequently do not possess the required properties or specificities to catalyze industrially useful transformations. This mini-review describes recent work using knowledge-guided site-specific mutagenesis of key active site residues to alter substrate specificity, stereospecificity and reaction specificity of these enzymes. In addition, examples of *de novo* designed enzymes that catalyze C-C bond reactions not found in nature will be discussed.

## Mini Review Article

## Introduction

Enzymes from the lyase (EC 4.1), ligase (EC 6.4) and hydrolase (EC 3.1, 3.7) classes have been explored for organic synthesis of value-added compounds and pharmaceuticals due to their ability to catalyze carbon-carbon bond formation (for reviews please see [1-4]). Natural enzymes however, do not exist for many industrially important transformations and therefore there is practical interest in tailoring the enzymes for novel functionalities. This can be accomplished through directed evolution approaches, involving random mutagenesis followed by screening of resultant enzyme variants for desired properties [5-8]. However, this approach tends to be labour intensive and the requirements for a screening method can limit the type of reactions that can be efficiently selected. Rational design approaches, involving site-specific mutagenesis of key residues of the enzyme is an alternative means to alter enzymes' properties or function. This often requires an intimate knowledge of the enzyme's structure and its relationship to function. In some cases, when functional significance of an amino acid is unknown, "semi-rational" approaches can be employed, targeting a specific amino acid for multiple mutations. There have been a number of recent examples of successful rational or semi-rational design of C-C bond forming enzymes, which will be the focus of this review. Some exciting developments on the *de novo* rational design of enzymes to catalyze novel C-C bond forming chemistries are also presented.

## Alterations to Substrate Specificity

One of the main limitations of enzymes as catalysts for synthetic chemistry is that natural enzymes can only transform a limited range of substrates. In the cases discussed below, modifications of enzymes' substrate binding sites could have profound effects on substrate specificity.

While many enzymes have evolved to accept phosphorylated substrates, synthetic chemists rarely want to prepare phosphorylated products. 2-deoxyribose-5-phosphate aldolase (DERA) has been modified to increase its preference for non-phosphorylated substrates. DERA is unique among aldolases as it catalyzes the reversible asymmetric aldol addition reaction of two aldehydes, acetaldehyde and D-glyceraldehyde-3-phosphate, to generate D-2-deoxyribose-5-phosphate [9]. Using the 1.05 Å crystal structure of DERA in complex with its natural substrate, two positively charged residues (K172 and R207) and three neutral active site residues (G205, S238, S239), located in the phosphate binding pocket were replaced with negatively charged aspartate or glutamate to change the enzyme specificity from the utilization of a negatively charged phosphorylated substrate to a non-phosphorylated neutral D-2-deoxyribose (DR) substrate. In each of the mutants, specificity constants ($k_{cat}/K_m$) for the natural substrate were reduced by 11-fold to $10^8$-fold. Specificity constants for DR were only improved in the S238D variant by 2.5-fold, possibly due to favorable hydrogen bonding interactions between the aspartate side chain carboxylate at position 238 and the terminal hydroxyl of D-glyceraldehyde. Interestingly, the S238D variant was able to catalyze a novel sequential aldol reaction using 3-azidopropinaldehyde, which is not a substrate for the wild-type enzyme, as the first acceptor and two molecules of acetaldehyde as donor to form an azidoethyl pyranos. This is a key intermediate in the synthesis of Lipitor™, a statin drug used for lowering blood cholesterol.

Specificity of an enzyme can also be altered by steric modifications to the active site of an enzyme. The substrate specificity of BphI, a class II pyruvate aldolase [10], was altered using the crystal structure of an orthologous aldolase, DmpG (PDB 1NVM), as a guide [11]. The natural substrate 4-hydroxy-2-oxopentanoate and 4-hydroxy-2-oxoacids containing unfunctionalized aldehydes of different lengths were modeled into the active site of DmpG [12] (Figure 1). Leu-87, which was proposed to be located proximal to the C4-methyl of 4-hydroxy-2-oxopentanoate and Leu-89, located at the opposite side of the active site from the pyruvyl moiety in BphI were replaced with smaller alanine residues. The replacement of Leu-87 with alanine reduced the specificity constants of the enzyme toward acetaldehyde by 32-fold while maintaining the same specificity constant for propionaldehyde as in the wild-type enzyme. The L89A variant exhibited similar $k_{cat}/K_m$ values for acetaldehyde and propionaldehyde as in the wild-type enzyme, however specificities toward butyraldehyde, pentaldehyde, and hexaldehyde increased

[a]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario Canada, N1G 2W1

* Corresponding author. Tel.: +1 5198244120; Fax: +1 5198371802
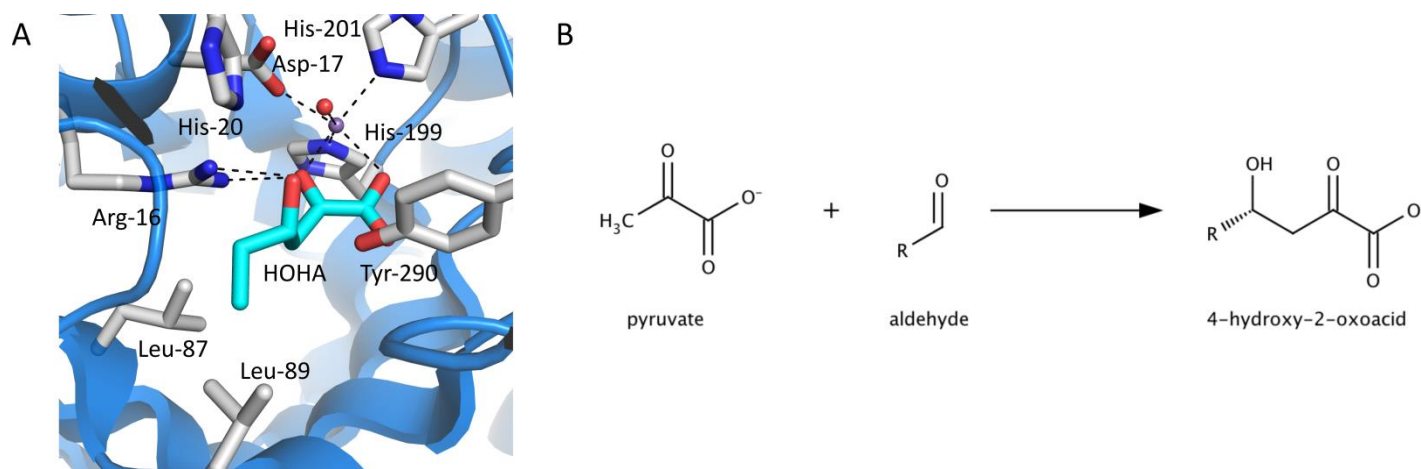E-mail address: sseah@uoguelph.ca (Stephen Y. K. Seah)

**Figure 1. Active site of pyruvate aldolase and the reaction it catalyzes.** (A) Active site of DmpG, an ortholog of BphI, containing the modeled substrate 4-hydroxy-2-oxohexanoate (HOHA). The alkyl chain of the 4-hydroxy-2-oxoacid extends towards Leu-89. Carbon atoms of the substrate are colored cyan, and the residues in the substrate binding sites are colored gray. Numbers correspond to residues in BphI. (B) The enzyme, BphI, catalyzes the aldol addition reaction between pyruvate and an aldehyde. R= CH3 and C5H11 for acetaldehyde and pentaldehyde, respectively.

dramatically. For example, the specificity of the L89A variant for pentaldehyde was ~50-fold higher than that of the wild type, making it a more efficient enzyme ($k_{cat}/K_m$ = 23.3 ± 2.8 M$^{-1}$ s$^{-1}$) for utilization of this longer chain aldehyde than the wild-type enzyme for its natural substrate, acetaldehyde ($k_{cat}/K_m$ = 13.4 ± 0.8 M$^{-1}$ s$^{-1}$). These results demonstrate that the alkyl moiety of the aldehyde substrate extends towards Leu-89, and reducing the side chain size of this residue can create an enzyme with increased specificity for longer chain aldehydes.

Dihydroxyacetone (DHA) and dihydroxyacetone phosphate (DHAP)-dependent aldolases have been extensively utilized in organic synthesis because they catalyze stereoselective aldol additions leading to the formation of useful chiral polyhydroxylated structures [13]. A number of studies have successfully utilized rational and semi-rational approaches, guided by X-ray crystal structures, to alter the substrate specificity through mutations in the phosphate-binding and other active site residues [14, 15].

Of special attention is FucA, a L-fuculose-1-phosphate aldolase from *E. coli*. DHAP aldolases, such as FucA, are specific for the DHAP donor substrate but can accept structurally diverse aldehyde acceptors. In spite of this diverse specificity towards the aldehyde acceptors, the enzyme does not effectively catalyze reactions involving N-Cbz-amino aldehydes (Cbz = benzyloxycarbonyl) containing α-branched substituents to produce polyhydroxylated pyrrolizidines glycosidase inhibitors. Using the crystal structure of FucA [16], Clapés and colleagues targeted Phe-131 and Phe-206 that compose a hydrophobic wall at the binding site for the natural acceptor substrate, L-lactaldehyde. Single alanine mutations of these residues were created to effectively remove these aromatic residues that sterically hinder the productive binding of branched α-substituted N-Cbz-amino aldehydes. While F131A and F206A variants exhibited a 5 to 100-fold reduction in specific activity for the natural reaction, the F131A was observed to catalyze the aldol addition reactions of DHAP to various N-Cbz-amino aldehyde derivatives 4 to 25-fold higher than that of wild-type FucA [17]. Moreover, F131A displayed aldol activity with (S)- and (R)-N-Cbz-prolinal, whereas no detectable activity was observed in wild-type FucA for the (R)-enantiomer. Computational binding models and molecular dynamic simulations further indicated that the phenyl group of Cbz can

occupy the space generated by the F131A and F206A and that in F131A, the Cbz-phenyl group is stabilized by a π-cation interaction with the charged ε-amine of Lys-205. This interaction would therefore increase binding affinity of these substrates relative to the wild-type enzyme. Thus, a single point mutation effectively eliminated a steric constraint, allowing for a non-natural π-cation interaction, allowing for the synthesis of novel polyhdroxylated pyrrolizidines of the hyacinthacine and alexine types.

Another such aldolase which has received considerable attention is the class I D-fructose-6-phosphate aldolase, FsaA from *E. coli*. Its catalytic potential arises from the fact that the enzyme can catalyze the asymmetric synthesis of polyhydroxylated compounds utilizing unphosphorylated dihydroxyacetone (DHA) instead of the more expensive DHAP. The aldolase catalyzes the reversible aldol addition of DHA, hydroxyacetone (HA) and hydroxybutanone (HB) with a variety of acceptor aldehydes. To improve the enzyme's specificity for DHA by allowing hydrogen bonding to occur with the C-1 hydroxyl of this substrate, an A129S variant was constructed. The $k_{cat}/K_m$ values for this variant towards DHA increased by 17-fold [18]. It is thought that the serine stabilizes the DHA donor substrate and may also be involved in the stabilization of the Schiff-base intermediate [19]. Despite the ability for the enzyme to utilize a number of donor substrates, the enzyme has limitations in its acceptor substrate tolerance. In an attempt to rationally design the enzyme to complete the aldol addition of DHA or HA to α-substituted aminoaldehydes, such as (S)- and (R)- N-Cbz-alaninals, Clapés and co-workers identified a putative residue in the active site thought to be involved in acceptor substrate-binding. Substitution of Ala-165 with glycine was envisaged to create more space in the active site to accommodate the C-α substituents of the aldehyde acceptors. Indeed, the variant catalyzed aldol addition reaction using HA and DHA as the donor and α-substituted aminoaldehydes as the acceptor with specific activities 4 to 175-fold higher than the wild-type enzyme [20]. Combining the two mutations, A129S and A165G, improved the activity of the enzyme in catalyzing the aldol addition of DHA and HA to selected aldehydes. These mutations did not result in changes in the stereochemical outcome of the reaction suggesting that the active site is not substantially altered.
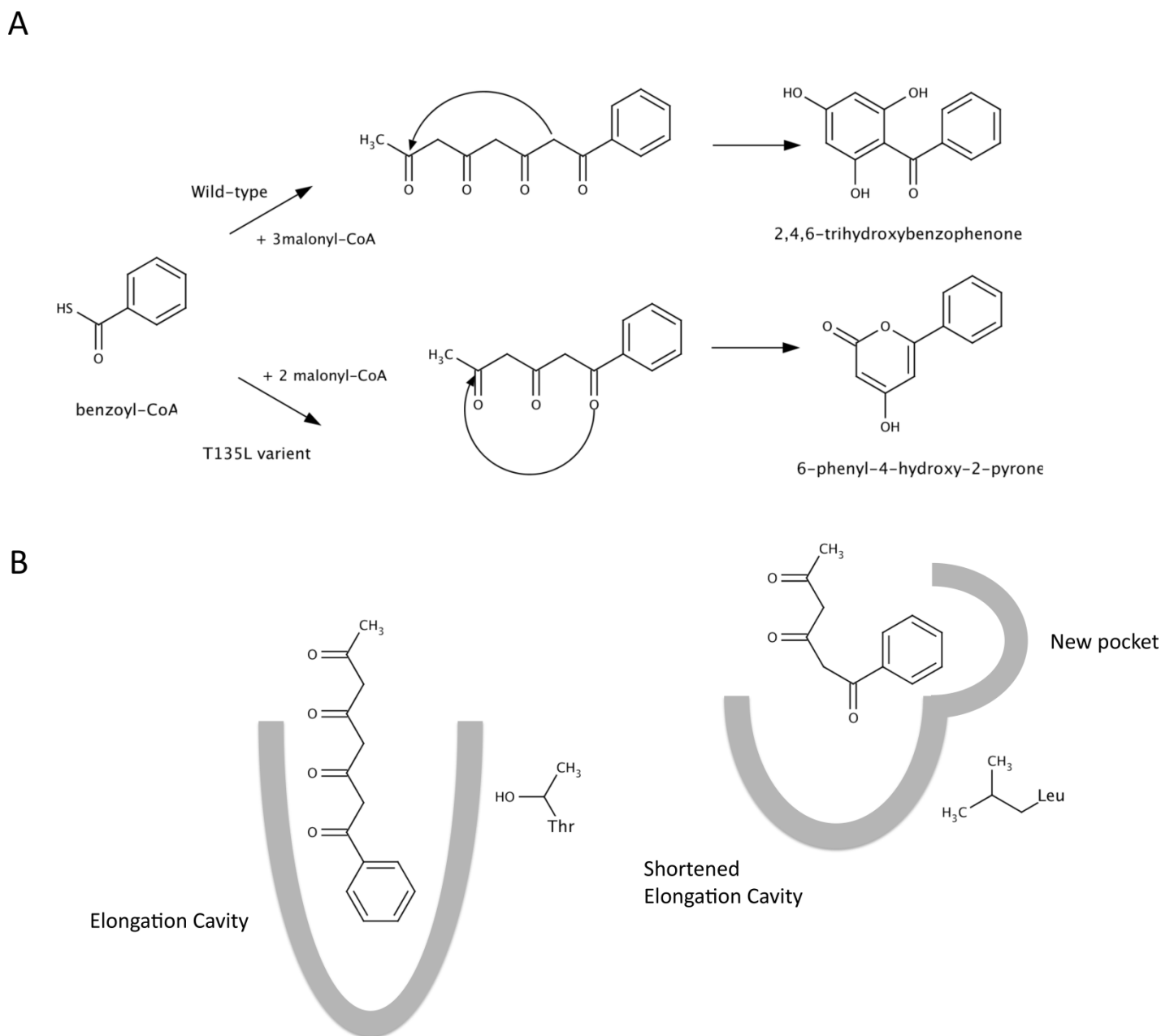
A



B



**Figure 2.** Postulated mechanism of wild-type BPS and the T135L variant. (A) The wild-type enzyme catalyzes the condensation of benzoyl-CoA with three malonyl-CoAs to form a tetraketide which cyclizes to 2,4,6-trihydroxybenzophenone. The T135L variant is only able to catalyze the condensation of benzoyl-CoA with two malonyl-CoAs, forming a triketide which cyclizes to 6-phenyl-4-hydroxy-2-pyrone. (B) Differences in active site structure result in altered product specificity. In the wild-type enzyme the benzene ring of benzoyl-CoA projects into the elongation tunnel as acetyl units from malonyl-CoA are added. This tunnel is open due to hydrogen bonding between Thr-135 and the backbone amides of Gln-165 and Gly-166. In the T135L variant, leucine projects into the tunnel, resulting in a steric block while eliminating the hydrogen-bonding with Gln-165 and Gly-166 and resulting in the formation of a new pocket. Favorable hydrophobic interactions between Leu-135 and the benzene ring of benzoyl-CoA result in a new orientation of substrate binding and elongation in the T135L variant.

Benzophenone derivatives constitute a class of secondary plant metabolites that include a number of pharmacologically active constituents, including garcinol and gambogic acid. The formation of the carbon skeleton of these benzophenone derivatives is catalyzed by benzophenone synthase (BPS), a type II polyketide synthase. In *Hypericum androsaemum,* BPS catalyzes the iterative condensation of benzoyl-CoA with three molecules of malonyl-CoA to give a linear tetraketide intermediate, which subsequently cyclizes into 2,4,6-trihydroxybenzophenone by Claisen condensation [21]. The enzyme also synthesizes small amounts of 6-phenyl-4-hydroxy-2-pyrone (phenylpyrone). Using molecular modeling based on the crystal structure of chalcone synthase (59.1% sequence identity with benzophenone synthase) (PDB 1CGZ) [22] as a guide, an attempt was made to alter substrate specificity. In the wild-type enzyme, Thr-135 forms hydrogen bonds with the backbone of Gly-166 (Figure 2). A T135L variant was created such that the leucine side chain protrudes into the elongation cavity and at the same time disrupting the hydrogen bond formed with Gly-166. This resulted in the opening of a new pocket within the active site to accommodate a phenyl group of the substrate. Relative to the wild-type enzyme, the variant catalyzed the addition of only two acetyl groups from malonyl-CoA to the benzoyl starter unit forming phenylpyrone. The
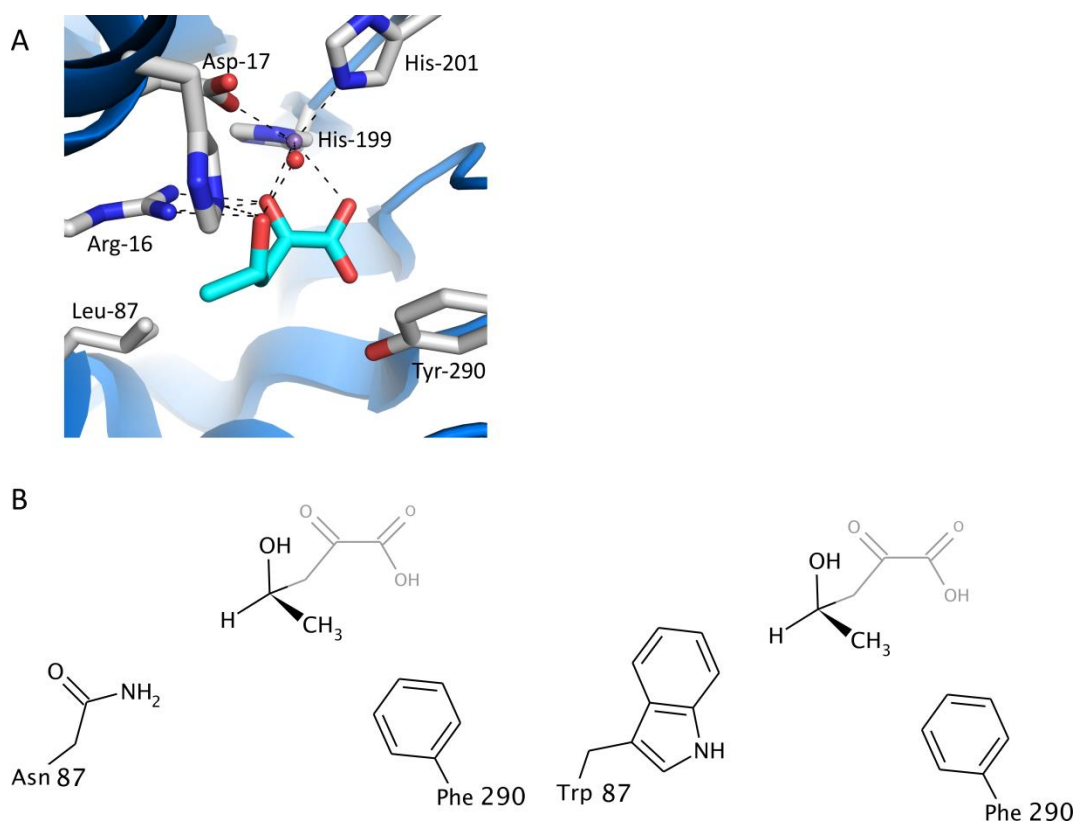
3

**Figure 3.** Engineering stereospecificity in BphI. (A) Model of 4(*S*)-hydroxy-2-oxopentanoate in the active site of DmpG, an ortholog of BphI. The chiral C4 carbon is positioned between Leu-87 and the phenolic oxygen of Tyr-290. Carbon atoms of the substrate are colored cyan, and the carbon atoms of residues in the substrate binding sites are colored gray. Numbers correspond to residues in BphI. (B) To allow for binding of the 4-methyl in the 4(R)-enantiomer, the phenolic oxygen was removed with a Y290F variant. At the same time, Leu87 was replaced with asparagine and tryptophan so that binding of the 4(S) enantiomer is disfavoured.

catalytic efficiency ($k_{cat}/K_m$) for the variant was almost unchanged relative to the wild-type enzyme as a 8-fold reduction in $k_{cat}$ was offset by a 8-fold reduction in $K_m$. When 3-hydroxybenzoyl-CoA is used as the starter substrate, the wild-type enzyme catalyzes the stepwise addition of three acetyl groups to give a tetraketide which cyclizes into 2,3',4,6-tetrahydroxybenzophenone. The T135L variant however is almost inactive with 3-hydroxy-benzoyl CoA as a starter substrate as the variant eliminates H-bonding between the enzyme and the 3-hydroxyl group of the aryl moeity. The 3-hydroxybenzoyl-primed triketide therefore becomes trapped in the new pocket of the T135L variant. Thus, the T135L substitution changes both the substrate and product specificity of BPS.

## Modifying enantiomeric selectivity

One of the main advantages of using enzymes as catalysts for C-C bond formation is their stereoselectivity, which is important for synthesis of pharmaceutical drugs. Examples of rational design of stereochemical control in C-C bond forming enzymes are discussed below.

The class II pyruvate aldolase BphI exhibits strict stereochemical control as it can only utilize or produce the (4*S*) enantiomer of 4-hydroxy-2-oxopentanoate [23]. Y290F and Y290S variants were found to lose stereochemical control and were able to catalyze aldol cleavage reactions of both (4*S*) and (4*R*) enantiomers of 4-hydroxy-2-oxopentanoate with similar catalytic efficiencies [12]. This is consistent with the proposal that the *p*-hydroxyl of Tyr-290 in the wild-type enzyme sterically restricts the binding of the (4*R*)-

enantiomer and its removal in the Y290F or Y290S variants enables the enzyme to utilize this enantiomer. Modeling of the 4-hydroxy-2-oxopentanoate in the active site of an orthologous enzyme placed the C4 chiral carbon between Leu-87 and Tyr-290 (Figure 3A) [12]. Replacement of Leu-87 with the polar asparagine (L87N) and the bulky tryptophan (L87W) residue were then attempted to create unfavorable polar-hydrophobic interactions or to introduce steric constraints to prevent binding of (4*S*) enantiomer. The double mutants, L87N/Y290F and L87W/Y290F were observed to catalyze the cleavage of (4*R*)-hydroxyl-2-oxopentanoate and not the (4*S*) enantiomer [24] (Figure 3B). In these double variants, pyruvate binding was unaffected and the enzyme maintained stereospecificity for substrates of various lengths that is opposite to that of the wild-type enzyme.

Stereospecificity of a class I aldolase, 2-keto-3-deoxygluconate aldolase, has also been improved by rational design [25]. This enzyme exhibits broad specificity for the aldol addition reaction between pyruvate and non-phosphorylated aldehydes [26]. The wild-type enzyme exhibits poor diasterocontrol in many of its aldol reactions, including the reaction of its natural substrates, pyruvate and D-glyceraldehyde producing a 55:45 racemate of D-2-keto-3-deoxygluconate (D-KDGlu) and D-keto-3deoxy-galactonate (D-KDGal) (Figure 4A). X-ray crystal structures of imine covalent complexes of this aldolase with pyruvate, D-KDGlu and D-KDGal have revealed that the carboxylate and C4-OH of D-KDGlu and D-KDGal are stabilized by common interactions. However, the C5-OH of D-KDGlu makes water mediated H-bonding interactions with the enzyme and the C6-OH is stabilized by direct H-bonding with
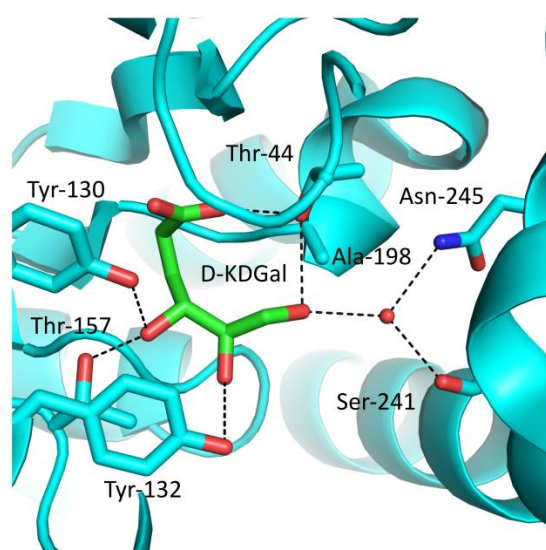
4
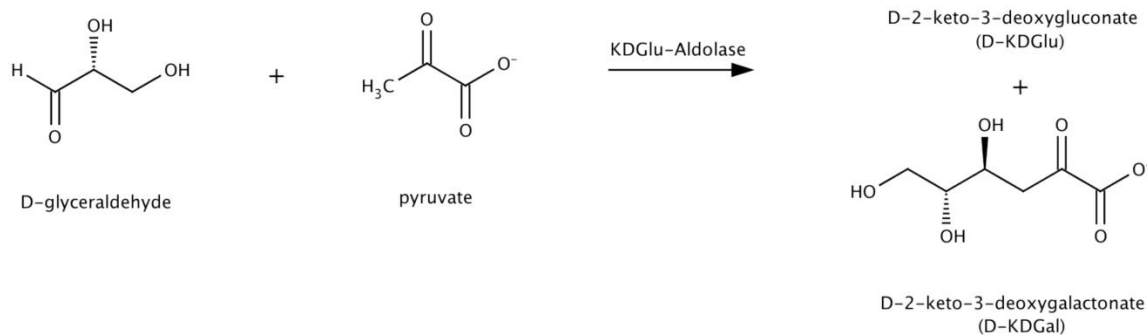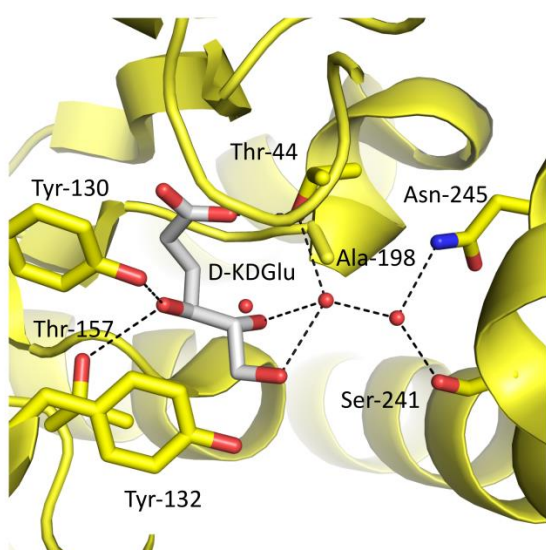
**Figure 4. Engineering stereochemical control in 2-keto-3-deoxygluconate aldolase.** (A) The wild-type aldolase catalyzes the nonstereoselective aldol addition reaction between pyruvate and D-glyceraldehyde. (B) X-ray structure of 2-keto-3-deoxygluconate aldolase from *Sulfolobus sofataricus* bound with D-KDGlu (PDB 1W3N). D-KDGlu makes water mediated H-bonding interactions with the enzyme and the C6-OH is stabilized by direct H-bonding with Tyr-132 and water mediated hydrogen bonding with Thr-44, Ser-241 and Asn-245 (C) X-ray structure of 2-keto-3-deoxygluconate aldolase from *Sulfolobus sofataricus* bound with D-KDGal (PDB 1W3T). The C5-OH of bound D-KDGal hydrogen bonds directly with Tyr-132 and the C6-OH interacts directly with Thr-44. Structures are depicted in cartoon with active site residues shown as sticks.

Tyr-132 and water mediated hydrogen bonding with Thr-44, Ser-241 and Asn-245 (Figure 4B). In comparison, the C5-OH of D-KDGal H-bonds directly with Tyr-132 and the C6-OH interacts directly with Thr-44 (Figure 4C). This indicates that binding of D-KDGlu is highly dependent on a conserved H-bond network composed of water molecules. Although Thr-157 plays a role in H-bonding to the C4-OH of both diastereoisomers, it is more important for binding of D-KDGal (distance between $C_\beta$-hydroxyl of the threonine side chain and C4-OH of D-KDGal and D-KDGlu is 2.47 Å and 3.07 Å, respectively). Thus, replacement of Thr-157 with residues of larger side chains, T157C and T157F, led to improved stereocontrol for D-KDGlu with 75% dr and 79% dr, respectively. To improve the stereochemical control for D-KDGlu, Tyr-132 that interacts directly with the C5-OH was replaced with valine. The Y132V variant catalyzed the selective formation of D-KDGlu in 79% dr. The Y132V/T157C and Y132V/T157F double variants were able to catalyze the formation of D-KDGlu with 91% dr and

93%, respectively. A separate set of variants was created to increase the selectivity of the enzyme for KDGal by reducing the H-bonding network required for stabilization of KDGlu. An A198L/T157G variant resulted in a small improvement in diastereocontrol, resulting in a preference for D-KDGal with a 72% dr. Asp-181, which is oriented towards Leu-198 was subsequently substituted with Gln to lengthen this polar amino acid and further direct the hydrophobic side chain of Leu-198 towards C5-OH of D-KDGlu, thus destabilizing the interaction with this substrate. In addition, the Leu-198 side chain is thought to stabilize the C1-C3 hydrophobic face of the D-glyceraldehyde conformer of D-KDGal. Indeed, the resulting triple variant A198L/T157G/D181Q was able to catalyze the diastereoselective formation of D-KDGal with 88 % dr. However, improved stereoselectivity came with a >70-fold decrease in $k_{cat}$.

Carboxmethylproline synthases (CMPSs) belong to the crotonase superfamily of enzymes that catalyze an early step in carbapenem antibiotic biosynthesis [27]. Under standard assay conditions the
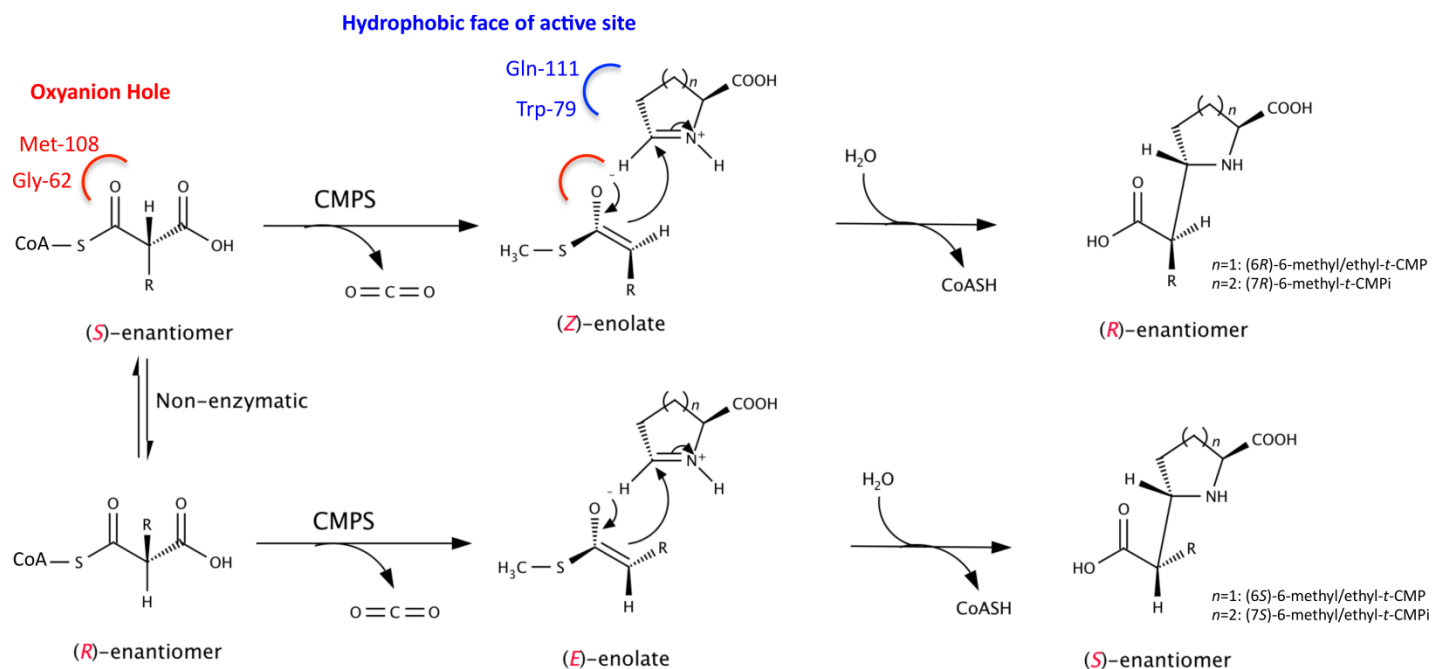
5

**Figure 5.** Proposed mechanism for the CMPS-catalyzed reaction. The decarboxylation of either alkylmalonyl-CoA enantiomer results in the formation of a either (*E*)- or (*Z*)-enolate intermediate, which reacts with an imine to give a specific diastereomer of carboxymethyl-substituted *N*-heterocycles.

CMPS from *Pectobacterium carotovorum*, known as CarB, catalyzes the reaction between L-GHP (an equilibrium mixture of L-glutamate semialdehyde, 5-hydroxyl-L-proline and L-pyrroline-5-carboxylate) with a C2 epimeric mixture of methylmalonyl-CoA to give a ~55:45 mixture of (6*R*)- and (6*S*)-epimers of 6-methyl carboxymethylproline (6-methyl-*t*-CMP) [28]. While no crystal structure of CarB in complex with L-GHP is present, previous knowledge that the enolate anion formed during the reaction is likely bound in an oxyanion hole composed of the backbone amides of Met-108 and Gly-62 led to targeting of these residues for site-specific mutagenesis in an attempt to introduce stereochemical control (Figure 5). Residues, Gln-111 and Trp-79 in CarB, thought to be involved in L-GHP binding, were also replaced by mutagenesis. Substitution of Met-108 in CarB with amino acid residues containing a β-branch, such as valine or isoleucine, resulted in an enzyme that preferentially produced the (6*R*) products (≥95:5 for 6*R*:6*S*) possibly due to steric clash between the β–methyl and reaction intermediates that result in the formation of the (6*S*) alkyl substituent product. Substitution of Trp-79 with alanine or phenylalanine on the other hand resulted in variants that preferentially produce the (6*S*) enantiomer product with ≥5:95 for 6*R*:6*S* [29]. It was observed that W79F and W79A variants also have a ~10-fold increase in catalytic rate relative to the wild-type enzyme. While the chemical basis behind the diastereoselectivity exhibited in the variants mentioned is not clear, it is anticipated that these substitutions may influence the binding of alkylmalonyl-CoA and the catalytic rates of the two half reactions.

## Engineering novel functionalities

Comparison of enzymes that share similarities in structure or reaction chemistry can be useful to engineer enzymes with novel functions.

Alanine racemase and L-threonine aldolase are pyridoxal 5'-phosphate (PLP)-dependent enzymes that are evolutionarily unrelated and have completely different tertiary and quaternary structures,

substrate preferences and reaction specificities [30]. However both enzymes form aldimine intermediates between PLP and their respective substrates, alanine or β-hydroxy-α-amino acid. In threonine aldolase from *Thermatoga maritima*, His-83 is thought to be the catalytic base catalyzing the proton abstraction from the β-hydroxyl of the substrate. Alanine racemace of *Geobacillus stearothermophilus* also contains an active site histidine (His-166). This residue forms a hydrogen bond with Tyr-265 and functions to lower its p$K_a$ [31] and act as a base required to convert L- to D-alanine [32]. Removal of the hydrogen bond interaction between His-166 and Tyr-265 in the single Y265A variant of alanine racemase, resulted in an enzyme that exhibits aldolase activity with an apparent $k_{cat}/K_m$ 2.3 × $10^5$-fold higher than racemase activity in the wild-type enzyme [33]. Racemase activity in the Y265A variant was reduced 4 × $10^5$-fold. The variant is also stereoselective, accepting only D-configured amino acids as substrates but displayed poor stereochemical control at the β carbon [34].

The enzyme 4-oxalocrontonate tautomerase (4-OT) catalyzes the conversion of 2-hydroxyhexa-2,4-dienedioate into 2-oxohex-3-enedioate in the degradation of aromatic hydrocarbons [35]. The enzyme belongs to the tautomerase superfamily that contain a unique catalytic amino-terminal proline. Recently it was discovered that the enzyme has a secondary activity in its ability to catalyze the aldol condensation between acetaldehyde and benzaldehyde with a $k_{cat}/K_m$ of 8.5 × $10^{-4}$ M$^{-1}$ s$^{-1}$ to yield cinnamaldehyde [36]. In the tautomerase reaction, Pro-1 of 4-OT acts as a nucleophile to form an enamine intermediate. In the presence of electrophiles, this nucleophilic intermediate can undergo carbon-carbon bond formation reactions. Therefore the promiscuous aldolase activity of OT is presumed to result from the cross-coupling of acetaldehyde and benzaldehyde to yield 3-hydroxy-3-phenylpropanal which subsequently undergoes dehydration to yield cinnamaldehyde. It has been proposed that the final hydrolysis step is rate limiting. In an attempt to increase this step in the reaction, Poelarends and co-workers used previous knowledge of a F50A variant that increased the accessibility of the active site to

the external aqueous environment without dramatic alterations to the p$K_a$ of Pro-I [37]. The resulting F50A variant exhibited a 600-fold increase in $k_{cat}/K_m$ in the aldol condensation reaction of acetaldehyde and benzaldehyde.

Tyrosine phenol-lyase (TPL) is a PLP-dependent enzyme that is structurally similar to aspartate aminotransferase [38]. TPL catalyzes the β-elimination of L-tyrosine to produce phenol, pyruvate and ammonium. In an attempt to switch the substrate specificity of TPL from L-tyrosine to dicarboxylic amino acids, homology modeling indicated that Val-283 in TPL occupies the same position as Arg-292 in aspartate aminotransferase, a residue that binds the carboxylate group in the aspartate side chain. Arg-100 in TPL on the other hand corresponds to Thr-I09 in aspartate aminotransferase, which interacts with the phosphate group of PLP. In the R100T/V283R double variant of TPL the β-elimination reaction towards dicarboxylic acids increase by at least $10^4$-fold compared to the wild-type enzyme and the rate of β-elimination using L-aspartate as substrate was 2-fold higher rate than that of L-tyrosine [39]. The rate of β-elimination was only one order of magnitude slower than that of L-tyrosine in the wild-type enzyme. Thus, the R100T/V283R variant was converted to a dicarboxylic amino acid β-lyase, an enzyme not found in nature without significant change to its reaction specificity.

The elimination of one catalytic function can also lead to the improvement of another secondary catalytic function. Pyruvate decarboxylase (PDC) is a thiamin diphosphate dependent enzyme that catalyzes the decarboxylation of pyruvate to form acetaldehyde and carbon dioxide in the presence of $Mg^{2+}$ (Figure 6) [40]. The enzyme is also able to catalyze a carboligation side reaction, whereby the carbanion/enamine intermediate, following decarboxylation of pyruvate, reacts with excess or exogenous aldehydes, to form 2-hydroxy ketones such as (R)-phenylacetyl carbinol, the precursor of several antiasthmatics such as ephedrine, pseudoephedrine, and norephedrine [41]. Kinetic and thermodynamic single-step analysis in conjunction with the X-ray crystal structure of PDC from *Zymomonas mobilis* implicated an active site residue Glu-473 in donating a proton to the 2-hydroxyethyl-ThDP carbanion/enamine intermediate to form acetaldehyde [42]. Consequently, in the E473Q variant, the protonation step occurs 2000-fold slower than the wild-type enzyme, and the catalytic cycle stalls at the carbanion/enamine intermediate state following pyruvate decarboxylation. Therefore this variant is more efficient at catalyzing formation of (R)-PAC in the presence of benzaldehyde [43]. The high enantioselectivity of the wild-type enzyme is retained in the variant.
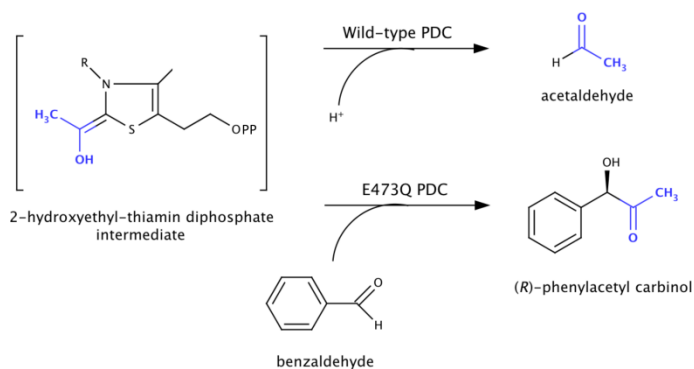


**Figure 6. Half reaction of pyruvate decarboxylase.** In the wild-type enzyme, protonation of the 2-hydroxyethyl-thiamin diphosphate intermediate leads to the release of acetaldehyde. In the E473Q variant, protonation occurs more slowly thus allowing for carboligation of benzaldehyde with the intermediate leading to the formation of (R)-phenylacetyl carbinol, a precursor of (-)-ephedrine.

## De novo rational computation design

The ultimate goal in rational protein design is the creation enzymes *de novo* i.e. creating synthetic protein catalysts. *De novo* design involves the selection of a catalytic mechanism for the desired reaction and modeling the reaction transition state(s). A protein scaffold is then designed to accommodate and stabilize the transition state. Usually several enzymes are produced, ranked and experimentally evaluated. C-C bond forming reactions are one of the best understood chemical reactions and not surprisingly some of the most successful examples of *de novo* designed enzymes are those that catalyze this reaction.

Aldolases of the Class I type that form a Schiff base intermediate with a catalytic lysine [9, 44] were designed to catalyze the carbon-carbon bond cleavage of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone, a compound not found in biological systems (Figure 7A) [45]. Four active site motifs with differing catalytic lysine environments, carbinolamine stabilization and proton abstraction were generated on five different protein scaffolds belonging to the triose phosphate-isomerase (TIM)-barrel and jelly-roll folds [46]. Of the 72 designs that were selected and experimentally characterized, 32 exhibited detectable aldolase activity. The most significant rate enhancement ($k_{cat}/k_{uncat}$ of $2 \times 10^4$) was accomplished through a catalytic lysine in a hydrophobic environment with a water molecule stabilizing the carbinolamine intermediate and catalyzing the proton abstraction of the C4-OH group of the substrate. Replacement of the putative catalytic lysine in all models dramatically decreased or abolished catalysis completely, suggesting that the observed activity was due to the formation of a schiff base intermediate with the substrate. Structures of two of the designed enzymes were solved, which showed that the active site residues superimpose well on the original design.

To date, no natural enzyme has been demonstrated to catalyze an intermolecular Diels-Alder reaction. Therefore *de-novo* rational design of an enzyme that catalyze a pericyclic [4 + 2] cycloaddition between a conjugated diene and a dienophile to form a substituted chiral cyclohexene compound, forming two carbon-carbon bonds and up to four new stereogenic centers in one step, was attempted. In the well-studied Diels-Alder reaction between 4-carboxybenzyl *trans*-I,3-butadiene-I carbamate and *N,N*-dimethylacrylamide, the dominant interaction in the transition state is the highest occupied molecular orbital (HOMO) of the diene, 4-carboxybenzyl *trans*-I,3-butadiene-I carbamate with the lowest unoccupied molecular orbital (LUMO) of the dienophile, *N,N*-dimethylacrylamide (Figure 7B). Thus, narrowing the energy gap between HOMO and LUMO is predicted to increase the rate of the reaction. An enzyme was devised whereby a hydrogen bond acceptor was positioned to interact with the carbamate NH of the diene, raising the energy of the HOMO energy while a hydrogen bond donor interacts with the carbonyl of the dienophile, lowering the LUMO energy. A binding pocket was also designed to position the substrates in the proper relative orientation by incorporating functional groups that can form hydrogen bonds with the substrates. Out of 207 stable protein scaffolds with backbone geometries that allow for orientation of the catalytic residues relative to the substrates, 84 designs were selected, 50 of which were soluble and can be purified [47]. Two enzymes with different scaffolds exhibited Diels-Alderase activity. The structure of one of these designs was solved using X-ray crystallography and showed agreement with the design model with a RMSD of 0.5 Å.

A challenge in *de novo* designed enzymes is that the protein backbone structure is not optimized for substrate binding or catalysis, resulting in low catalytic efficiencies. Directed evolution approaches are one way to circumvent the need for precise backbone modeling to enhance
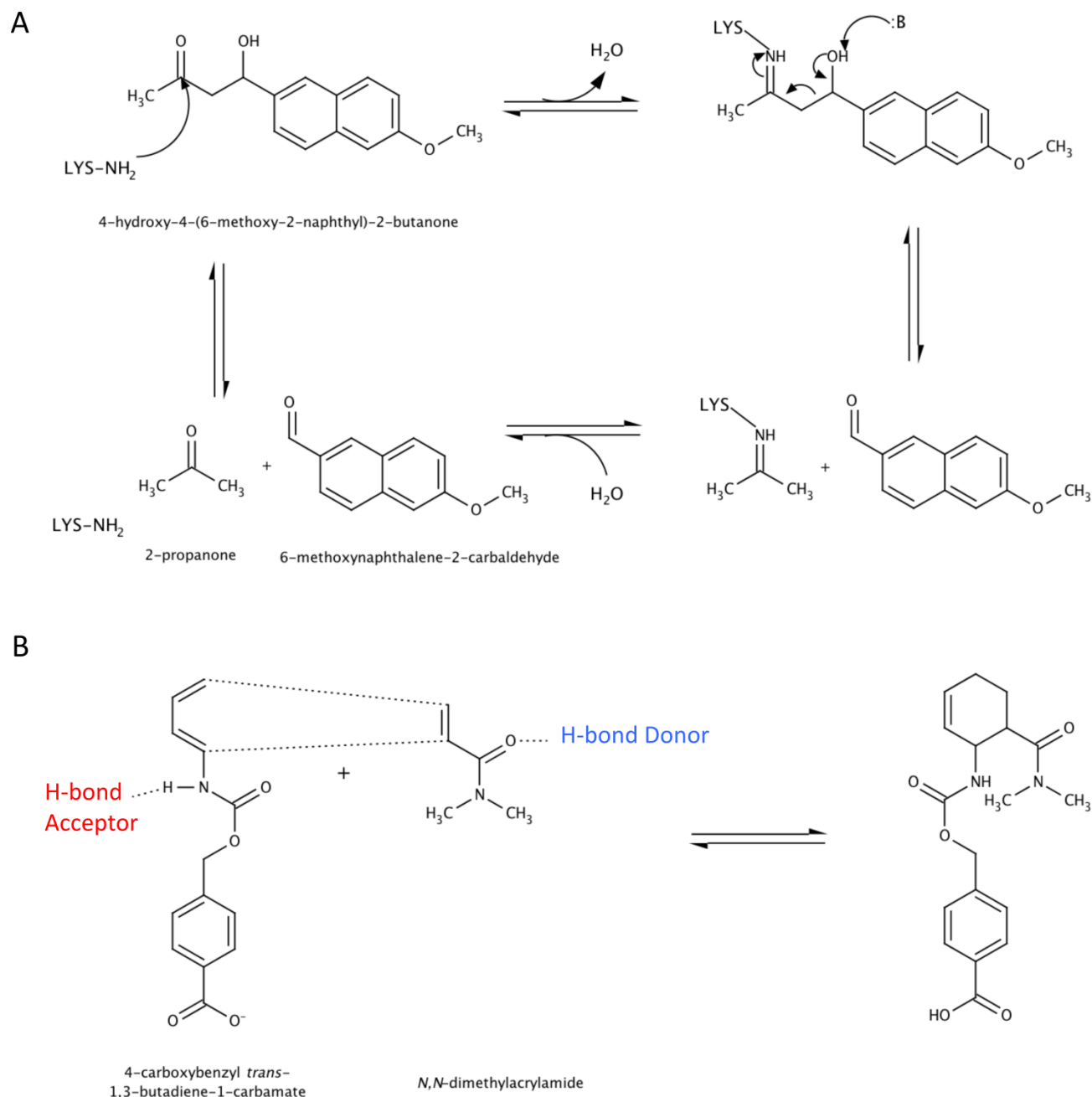
**Figure 7. Chemical reactions catalyzed by de novo designed enzymes.** (A) The retro-aldol reaction involves the formation of a Schiff base intermediate between a catalytic lysine and the substrate 4-hdyroxy-4-(6-methyoxy-2naphthyl)-2-butanone. C-C bond cleavage occurs following proton abstraction at the C4-OH of the substrate forming 2-propanone and 6-methoxynaphthalene-2-carbaldehyde. (B) The Diels-Alder reaction occurs between a dienophile I (4-carboxy *trans* 1,3-butadiene-1-carbamate) and diene II (*N,N*-dimethylacrylamide) which undergoes a pericyclic [4+2] cycloaddition to form a chiral cyclohexene ring.

the catalytic efficiency of designed enzymes. An alternative approach is to enlist the aid of the internet community to perform time consuming human interpretation and judgment required for modeling of protein backbone structure. This has been successfully implemented in the crowdsourcing program foldit. Users (players) designed enzyme was found to improve the catalytic efficiency of the diels alderase mentioned above by up to 19-fold [48].

knowledge based, rational approaches. Ultimately, such studies also strengthen our understanding of protein structure-function relationships that will in turn advance our ability to engineer proteins with desired functions with greater precision.

## Summary

The examples discussed above demonstrate that it is possible to alter the function carbon-carbon bond forming enzymes using

## Acknowledgements

## References

1. Clapes, P., Fessner, W. D., Sprenger, G. A., and Samland, A. K. (2010) Recent progress in stereoselective synthesis with aldolases, Curr Opin Chem Biol 14: 154-167.
2. Brovetto, M., Gamenara, D., Mendez, P. S., and Seoane, G. A. (2011) C-C bond-forming lyases in organic synthesis, Chem Rev 111: 4346-4403.
3. Clapés, P., and Garrabou, X. (2011) Current Trends in Asymmetric Synthesis with Aldolases, Advanced Synthesis & Catalysis 353: 2263-2283.
4. Dean, S. M., Greenberg, W. A., and Wong, C.-H. (2007) Recent Advances in Aldolase-Catalyzed Asymmetric Synthesis, Advanced Synthesis & Catalysis 349: 1308-1320.
5. Bolt, A., Berry, A., and Nelson, A. (2008) Directed evolution of aldolases for exploitation in synthetic organic chemistry, Arch Biochem Biophys 474: 318-330.
6. Jaeger, K. E., Eggert, T., Eipper, A., and Reetz, M. T. (2001) Directed evolution and the creation of enantioselective biocatalysts, Appl Microbiol Biotechnol 55: 519-530.
7. Fong, S., Machajewski, T. D., Mak, C. C., and Wong, C. (2000) Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars, Chem Biol 7: 873-883.
8. Wada, M., Hsu, C. C., Franke, D., Mitchell, M., Heine, A., Wilson, I., and Wong, C. H. (2003) Directed evolution of N-acetylneuraminic acid aldolase to catalyze enantiomeric aldol reactions, Bioorg Med Chem 11: 2091-2098.
9. Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H., and Wilson, I. A. (2001) Observation of covalent intermediates in an enzyme mechanism at atomic resolution, Science 294: 369-374.
10. Baker, P., Pan, D., Carere, J., Rossi, A., Wang, W., and Seah, S. Y. (2009) Characterization of an aldolase-dehydrogenase complex that exhibits substrate channeling in the polychlorinated biphenyls degradation pathway, Biochemistry 48: 6551-6558.
11. Manjasetty, B. A., Powlowski, J., and Vrielink, A. (2003) Crystal structure of a bifunctional aldolase-dehydrogenase: sequestering a reactive and volatile intermediate, Proc Natl Acad Sci U S A 100: 6992-6997.
12. Baker, P., Carere, J., and Seah, S. Y. (2011) Probing the Molecular Basis of Substrate Specificity, Stereospecificity, and Catalysis in the Class II Pyruvate Aldolase, BphI, Biochemistry 50: 3559-3569.
13. Calveras, J., Egido-Gabas, M., Gomez, L., Casas, J., Parella, T., Joglar, J., Bujons, J., and Clapes, P. (2009) Dihydroxyacetone phosphate aldolase catalyzed synthesis of structurally diverse polyhydroxylated pyrrolidine derivatives and evaluation of their glycosidase inhibitory properties, Chemistry 15: 7310-7328.
14. Garrabou, X., Joglar, J., Parella, T., Crehuet, R., Bujons, J., and Clapés, P. (2011) Redesign of the Phosphate Binding Site of L-Rhamnulose- 1-Phosphate Aldolase towards a Dihydroxyacetone Dependent Aldolase, Advanced Synthesis & Catalysis 353: 89-99.
15. Schneider, S., Gutierrez, M., Sandalova, T., Schneider, G., Clapes, P., Sprenger, G. A., and Samland, A. K. (2010) Redesigning the active site of transaldolase TalB from Escherichia coli: new variants with improved affinity towards nonphosphorylated substrates, Chembiochem 11: 681-690.
16. Joerger, A. C., Gosse, C., Fessner, W. D., and Schulz, G. E. (2000) Catalytic action of fuculose 1-phosphate aldolase (class II) as derived from structure-directed mutagenesis, Biochemistry 39: 6033-6041.
17. Garrabou, X., Gomez, L., Joglar, J., Gil, S., Parella, T., Bujons, J., and Clapes, P. (2010) Structure-guided minimalist redesign of the L-fuculose-1-phosphate aldolase active site: expedient synthesis of novel polyhydroxylated pyrrolizidines and their inhibitory properties against glycosidases and intestinal disaccharidases, Chemistry 16: 10691-10706.
18. Castillo, J. A., Guérard-Hélaine, C., Gutiérrez, M., Garrabou, X., Sancelme, M., Schürmann, M., Inoue, T., Hélaine, V., Charmantray, F., Gefflaut, T., Hecquet, L., Joglar, J., Clapés, P., Sprenger, G. A., and Lemaire, M. (2010) A Mutant D-Fructose-6-Phosphate Aldolase (Ala129Ser) with Improved Affinity towards Dihydroxyacetone for the Synthesis of Polyhydroxylated Compounds, Advanced Synthesis & Catalysis 352: 1039-1046.
19. Thorell, S., Schurmann, M., Sprenger, G. A., and Schneider, G. (2002) Crystal structure of decameric fructose-6-phosphate aldolase from Escherichia coli reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family, J Mol Biol 319: 161-171.
20. Gutierrez, M., Parella, T., Joglar, J., Bujons, J., and Clapes, P. (2011) Structure-guided redesign of D-fructose-6-phosphate aldolase from E. coli: remarkable activity and selectivity towards acceptor substrates by two-point mutation, Chem Commun (Camb) 47: 5762-5764.
21. Liu, B., Falkenstein-Paul, H., Schmidt, W., and Beerhues, L. (2003) Benzophenone synthase and chalcone synthase from Hypericum androsaemum cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases, Plant J 34: 847-855.
22. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis, Nat Struct Biol 6: 775-784.
23. Wang, W., Baker, P., and Seah, S. Y. (2010) Comparison of two metal-dependent pyruvate aldolases related by convergent evolution: substrate specificity, kinetic mechanism, and substrate channeling, Biochemistry 49: 3774-3782.
24. Baker, P., and Seah, S. Y. (2012) Rational design of stereoselectivity in the class II pyruvate aldolase BphI, J Am Chem Soc 134: 507-513.
25. Royer, S. F., Haslett, L., Crennell, S. J., Hough, D. W., Danson, M. J., and Bull, S. D. (2010) Structurally informed site-directed mutagenesis of a stereochemically promiscuous aldolase to afford stereochemically complementary biocatalysts, J Am Chem Soc 132: 11753-11758.
26. Wolterink-van Loo, S., van Eerde, A., Siemerink, M. A., Akerboom, J., Dijkstra, B. W., and van der Oost, J. (2007) Biochemical and structural exploration of the catalytic capacity of Sulfolobus KDG aldolases, Biochem J 403: 421-430.
27. Sleeman, M. C., and Schofield, C. J. (2004) Carboxymethylproline synthase (CarB), an unusual carbon-carbon bond-forming enzyme of the crotonase superfamily involved in carbapenem biosynthesis, J Biol Chem 279: 6730-6736.
28. Hamed, R. B., Batchelar, E. T., Mecinovic, J., Claridge, T. D., and Schofield, C. J. (2009) Evidence that thienamycin biosynthesis proceeds via C-5 epimerization: ThnE catalyzes the formation of (2S,5S)-trans-carboxymethylproline, Chembiochem 10: 246-250.

29. Hamed, R. B., Gomez-Castellanos, J. R., Thalhammer, A., Harding, D., Ducho, C., Claridge, T. D., and Schofield, C. J. (2011) Stereoselective C-C bond formation catalysed by engineered carboxymethylproline synthases, Nat Chem 3: 365-371.

30. Kielkopf, C. L., and Burley, S. K. (2002) X-ray structures of threonine aldolase complexes: structural basis of substrate recognition, Biochemistry 41: 11711-11720.

31. Sun, S., and Toney, M. D. (1999) Evidence for a two-base mechanism involving tyrosine-265 from arginine-219 mutants of alanine racemase, Biochemistry 38: 4058-4065.

32. Watanabe, A., Yoshimura, T., Mikami, B., and Esaki, N. (1999) Tyrosine 265 of alanine racemase serves as a base abstracting alpha-hydrogen from L-alanine: the counterpart residue to lysine 39 specific to D-alanine, J Biochem 126: 781-786.

33. Seebeck, F. P., and Hilvert, D. (2003) Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation, J Am Chem Soc 125: 10158-10159.

34. Seebeck, F. P., Guainazzi, A., Amoreira, C., Baldridge, K. K., and Hilvert, D. (2006) Stereoselectivity and expanded substrate scope of an engineered PLP-dependent aldolase, Angew Chem Int Ed Engl 45: 6824-6826.

35. Harayama, S., Rekik, M., Ngai, K. L., and Ornston, L. N. (1989) Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via meta cleavage in Pseudomonas putida, J Bacteriol 171: 6251-6258.

36. Zandvoort, E., Baas, B. J., Quax, W. J., and Poelarends, G. J. (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity, Chembiochem 12: 602-609.

37. Czerwinski, R. M., Harris, T. K., Massiah, M. A., Mildvan, A. S., and Whitman, C. P. (2001) The structural basis for the perturbed pKa of the catalytic base in 4-oxalocrotonate tautomerase: kinetic and structural effects of mutations of Phe-50, Biochemistry 40: 1984-1995.

38. Antson, A. A., Strokopytov, B. V., Murshudov, G. N., Isupov, M. N., Harutyunyan, E. H., Demidkina, T. V., Vassylyev, D. G., Dauter, Z., Terry, H., and Wilson, K. S. (1992) The polypeptide chain fold in tyrosine phenol-lyase, a pyridoxal-5'-phosphate-dependent enzyme, FEBS Lett 302: 256-260.

39. Mouratou, B., Kasper, P., Gehring, H., and Christen, P. (1999) Conversion of tyrosine phenol-lyase to dicarboxylic amino acid beta-lyase, an enzyme not found in nature, J Biol Chem 274: 1320-1325.

40. Candy, J. M., and Duggleby, R. G. (1998) Structure and properties of pyruvate decarboxylase and site-directed mutagenesis of the Zymomonas mobilis enzyme, Biochim Biophys Acta 1385: 323-338.

41. Shin, H. S., and Rogers, P. L. (1996) Production of L-phenylacetylcarbinol (L-PAC) from benzaldehyde using partially purified pyruvate decarboxylase (PDC), Biotechnol Bioeng 49: 52-62.

42. Meyer, D., Neumann, P., Parthier, C., Friedemann, R., Nemeria, N., Jordan, F., and Tittmann, K. (2010) Double duty for a conserved glutamate in pyruvate decarboxylase: evidence of the participation in stereoelectronically controlled decarboxylation and in protonation of the nascent carbanion/enamine intermediate, Biochemistry 49: 8197-8212.

43. Meyer, D., Walter, L., Kolter, G., Pohl, M., Muller, M., and Tittmann, K. (2011) Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential, J Am Chem Soc 133: 3609-3616.

44. Fullerton, S. W., Griffiths, J. S., Merkel, A. B., Cheriyan, M., Wymer, N. J., Hutchins, M. J., Fierke, C. A., Toone, E. J., and Naismith, J. H. (2006) Mechanism of the Class I KDPG aldolase, Bioorg Med Chem 14: 3002-3010.

45. Tanaka, F., Fuller, R., Shim, H., Lerner, R. A., and Barbas, C. F., 3rd. (2004) Evolution of aldolase antibodies in vitro: correlation of catalytic activity and reaction-based selection, J Mol Biol 335: 1007-1018.

46. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes, Science 319: 1387-1391.

47. Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction, Science 329: 309-313.

48. Eiben, C. B., Siegel, J. B., Bale, J. B., Cooper, S., Khatib, F., Shen, B. W., Players, F., Stoddard, B. L., Popovic, Z., and Baker, D. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players, Nat Biotechnol 30: 190-192.

**What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?**

- Easy 5 step online submission system & online manuscript tracking
- Fastest turnaround time with thorough peer review
- Inclusion in scholarly databases
- Low Article Processing Charges
- Author Copyright
- Open access, available to anyone in the world to download for free

**WWW.CSBJ.ORG**