BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Enzymes in chemical transformations

Edited by Karen N. Allen, Christian P. Whitman and Donald Hilvert

Imprint

Beilstein Journal of Organic Chemistry
www.bjoc.org
ISSN 1860-5397
Email: journals-support@beilstein-institut.de

The *Beilstein Journal of Organic Chemistry* is published by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

Beilstein-Institut zur Förderung der Chemischen Wissenschaften
Trakehner Straße 7–9
60487 Frankfurt am Main
Germany
www.beilstein-institut.de

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

The enzymes of microbial nicotine metabolism

Paul F. Fitzpatrick

Review

Open Access

Address:
Department of Biochemistry and Structural Biology, University of Texas Health Science Center, San Antonio, TX, 78229, USA

Email:
Paul F. Fitzpatrick - fitzpatrickp@uthscsa.edu

Keywords:
biodegradation; enzyme mechanism; flavoprotein; metabolic pathway; nicotine

Beilstein J. Org. Chem. **2018,** *14,* 2295–2307.
doi:10.3762/bjoc.14.204

Received: 14 May 2018
Accepted: 20 August 2018
Published: 31 August 2018

This article is part of the thematic issue "Enzymes in chemical transformations".

Guest Editor: K. N. Allen

Abstract

Because of nicotine's toxicity and the high levels found in tobacco and in the waste from tobacco processing, there is a great deal of interest in identifying bacteria capable of degrading it. A number of microbial pathways have been identified for nicotine degradation. The first and best-understood is the pyridine pathway, best characterized for *Arthrobacter nicotinovorans*, in which the first reaction is hydroxylation of the pyridine ring. The pyrrolidine pathway, which begins with oxidation of a carbon–nitrogen bond in the pyrrolidine ring, was subsequently characterized in a number of pseudomonads. Most recently, a hybrid pathway has been described, which incorporates the early steps in the pyridine pathway and ends with steps in the pyrrolidine pathway. This review summarizes the present status of our understanding of these pathways, focusing on what is known about the individual enzymes involved.

Introduction

The toxic alkaloid (*S*)-nicotine (L-nicotine) is found at high levels in tobacco leaves and the waste from tobacco processing. The resulting interest in developing environmentally friendly methods of degrading nicotine has driven studies of microbial pathways for metabolizing the compound, with the possible additional benefit of using the enzymes involved to synthesize specialty chemicals [1,2]. To date, the best-characterized bacterial pathways are those of *Arthrobacter nicotinovorans* and several pseudomonads. These are, respectively, known as the pyridine and pyrrolidine pathways due to the initial reactions in each. More recently, additional pathways have been described that combine steps from the pyridine and pyrrolidine pathways. To a large extent the descriptions of this metabolism have focused on the genes involved. An exception to this is the review by Brandsch [3], which describes the *Arthrobacter* pathway at a biochemical level. In the more than a decade since that review was published, a great deal has been learned about other pathways for nicotine metabolism and the enzymes involved.

The goal of the present report is to summarize our present understanding of the different pathways by which microbes metabolize nicotine, focusing on the enzymes.

Review

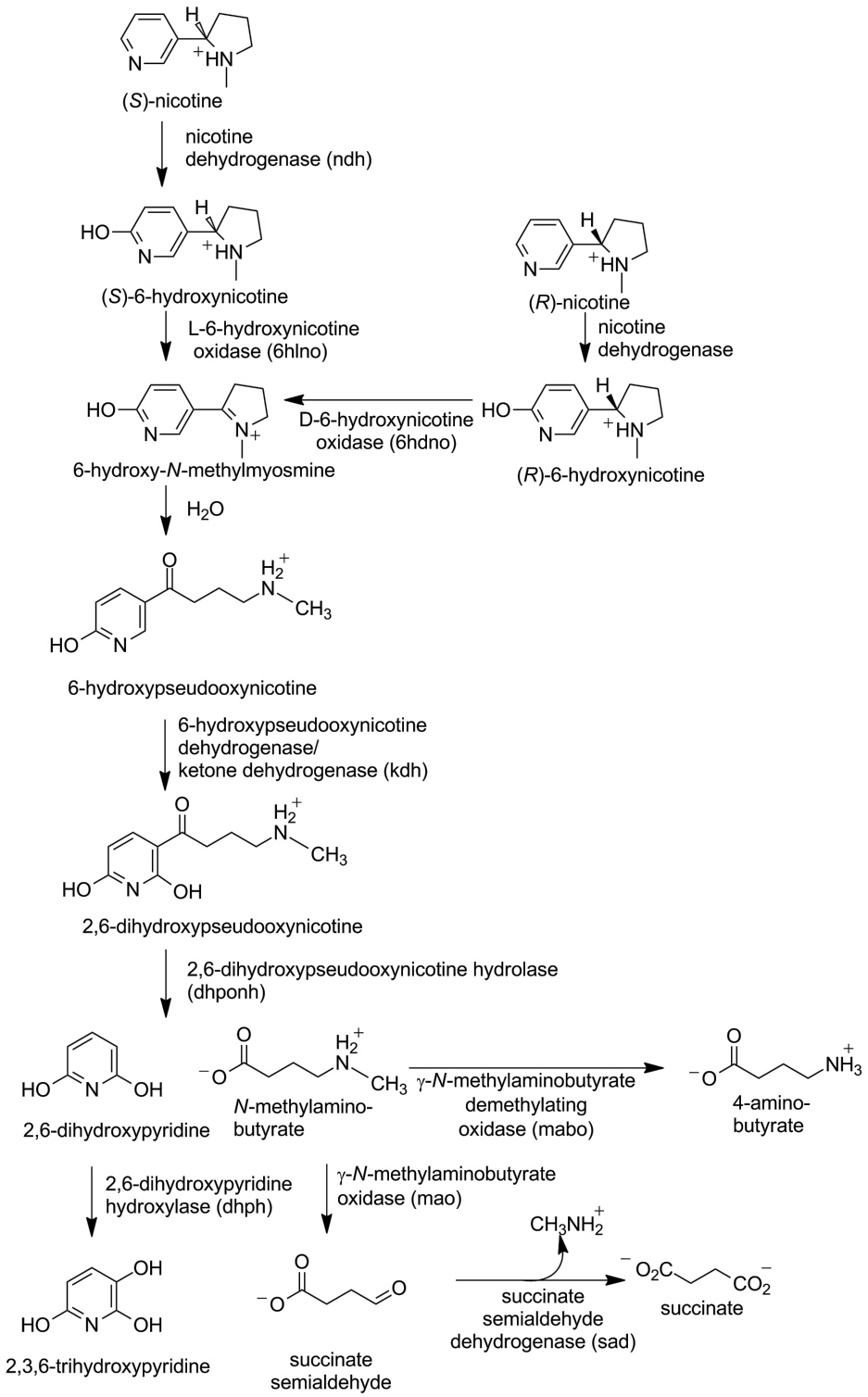
The pyridine pathway

In *A. nicotinovorans* the enzymes involved in nicotine metabolism are found on the plasmid pAO1 [4], and the sequencing of this plasmid was a major step in elucidating the pathway [5]. A similar pathway has been described for *Nocardioides* sp. JS614; in this case the genes are chromosomal [6]. As shown in Scheme 1, the pathway begins with hydroxylation of the pyridyl ring of nicotine by the enzyme nicotine dehydrogenase to yield 6-hydroxynicotine [7]. Based on the gene sequence, this enzyme was identified as a member of the family of molybdopterin enzymes that also includes xanthine oxidoreductase and aldehyde oxidase [8]. Comparison of the pAO1 sequence with that of xanthine oxidoreductase identified *ndhs*, *ndhm*, and *ndhl* (initially designated *ndhABC*) as coding for three proteins: a 14.9 kDa subunit containing an iron–sulfur cluster, a 30 kDa subunit with an FAD binding site, and an 87.7 kDa subunit containing the molybdopterin site, respectively [9,10]. Consistent with this identification, expression of the active enzyme required molybdopterin [9], and pAO1 contains a number of genes that have been identified as coding for proteins involved in uptake of molybdenum and biosynthesis of the molybdopterin cofactor [11]. The mechanism of Scheme 2 can be written for nicotine dehydrogenase by analogy to the mechanism of xanthine oxidoreductase [8]. Here, the oxygen that is incorporated into the product initially comes from water, and the two electrons produced are transferred through the iron–sulfur subunit to the FAD and thence to the final electron acceptor.

L-6-Hydroxynicotine oxidase (LHNO) catalyzes the subsequent oxidation of L-6-hydroxynicotine to 6-hydroxy-*N*-methylmysomine [12]. Purified LHNO contains non-covalently bound FAD [13], and the gene sequence is most similar to those of eukaryotic monoamine oxidases [14]. Several high-resolution structures of the enzyme from *A. nicotinovorans* are available, including substrate and product complexes [15]. These structures confirm that the protein is a member of the monoamine oxidase (MAO) family of flavoproteins (Figure 1) [16]. The reaction product was originally identified as arising from oxidation of the C2–C3 bond of the pyrrolidine ring [17]. Based on the structures and this product identification, a detailed mechanism was proposed in which initial oxidation of L-6-hydroxynicotine in the active site is followed by hydrolysis of the oxidized amine in a second site to yield 6-hydroxypseudooxynicotine (Scheme 3) [15]. However, a recent analysis of the structure of the product of the LHNO reaction utilizing NMR and continuous-flow mass spectrometry established that the enzyme catalyzes oxidation of the C2–N bond, not the C2–C3 bond, in line with the typical reactions catalyzed by members of the MAO family [18]. In addition, mutagenesis of His187, Glu300, and Tyr407 established that they are not involved in catalysis. Subsequent mechanistic studies of the reaction using pH and solvent isotope effects established that the reaction catalyzed by LHNO is the same as other flavin amine oxidases, direct hydride transfer from the uncharged amine to the flavin (Scheme 4) [19,20]. Hydrolysis to form 6-hydroxypseudooxynicotine occurs in solution after release of the oxidized amine from the enzyme.

While the dominant form of nicotine found in tobacco is (*S*)-nicotine, the (*R*)-stereoisomer is also found at detectable levels [22]. Nicotine dehydrogenase is reported not to be stereospecific, in that it can catalyze the hydroxylation of (*R*)-nicotine to (*R*)-6-hydroxynicotine; thus, this enzyme is a likely candidate for the enzyme catalyzing the first step in the metabolism of both stereoisomers [23]. The subsequent step requires an additional enzyme. The pAO1 plasmid contains the gene for a D-6-hydroxynicotine oxidase (DHNO) in addition to that for LHNO. The product of the reaction catalyzed by DHNO is identical to that of the LHNO reaction, so that this enzyme was also initially identified as catalyzing the oxidation of the C2–C3 bond [17]. However, NMR analysis of the product has also recently established that DHNO catalyzes oxidation of the C2–N bond [24]. The sequence of DHNO from *A. nicotinovorans* identifies it as a member of the p-cresol methylhydroxylase/vanillyl oxidase family of flavoproteins [25]. As is common for members of this family, the FAD in DHNO is covalently bound to the protein, in this case through a C8α-histidyl linkage [26]. The subsequent determination of the crystal structure of the enzyme confirmed these conclusions (Figure 2) [27]. Docking of (*R*)-6-hydroxynicotine into the structure yielded a model for substrate binding.

Vanillyl oxidase catalyzes the oxidation of 4-hydroxybenzyl alcohols, the oxidative deamination of 4-hydroxybenzylamines, and the oxidative demethylation of 4-(methoxymethyl)phenols via a quinone methide intermediate [28,29]. Based on this precedent and the assumption that DHNO oxidizes the C2–C3 bond, Koetter and Schultz [27] proposed the mechanism shown in Scheme 5 for DHNO. However, members of the *p*-cresol methylhydroxylase/vanillyl oxidase family catalyze an extremely diverse set of reactions, including oxidation of nonaromatic alcohols and amines [30], and DALI [31] identifies several enzymes catalyzing oxidation of nonaromatic substrates as having similar structures to DHNO. Indeed, recent mechanistic studies of DHNO are more consistent with the simple mechanism of Scheme 4 (Fitzpatrick et al., manuscript in preparation). The proposed quinone methide is not detected during



Scheme 1: Nicotine catabolism in *A. nicotinovorans*. The respective gene names are given in parentheses.

stopped-flow analyses of either the wild-type enzyme or the E352Q variant, (*R*)-6-chloronicotine and (*R*)-nicotine, which would not form the quinone methide, are still substrates, and there is no solvent isotope effect on amine oxidation. In addition, DHNO E350L/E352D has been developed as a reagent for stereospecific oxidation of a variety of (*R*)-amines, including a

Scheme 2: Hydroxylation of nicotine by the molybdopterin cofactor of nicotine dehydrogenase.

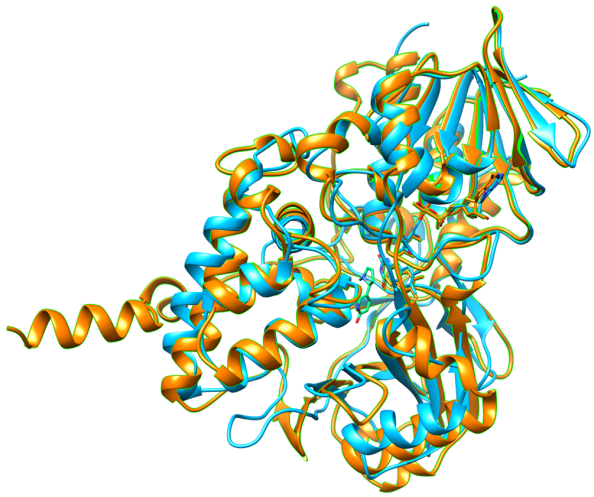


Figure 1: Overlay of the structure of LHNO (blue, pdb file 3NG7) with that of human MAO B (orange, pdb file 2FXU). The bound 6-hydroxynicotine is shown with green carbons.

nicotine analog that does not contain an aromatic ring [24]. These results provide further evidence against the mechanism shown in Scheme 5 for DHNO.

Water reacts with the 6-hydroxy-*N*-methylmyosmine formed by either LHNO or DHNO to form 6-hydroxypseudooxynicotine in a reaction that appears to be non-enzymatic. 6-Hydroxypseudooxynicotine dehydrogenase (also known as ketone dehydrogenase [3]) then catalyzes the hydroxylation of the pyridyl ring of 6-hydroxypseudooxynicotine to form 2,6-dihydroxypseudooxynicotine [32]. Based on the sequence of pAO1, the enzyme was identified as a molybdopterin enzyme containing three subunits coded for by the *kdha*, *dhb* and *kdhc* genes [14]. The predicted sequences of Kdha and Kdhb show significant similarity to the small and medium subunits of nicotine dehydrogenase, while that of Kdhc shows the highest similarity to chicken xanthine dehydrogenase. The spectral properties of the partially purified protein are consistent with 6-hydroxypseudooxynicotine dehydrogenase being a molybdopterin protein, and the recombinant Kdhc (also known as KdhL) contains Mo and a cofactor derived from CTP [33]. While no mechanistic studies of the enzyme have been reported, its mechanism is likely to resemble those of nicotine dehydrogenase (Scheme 2) and other molybdopterin enzymes [8].

Scheme 4: Mechanism of LHNO.

Scheme 3: Proposed mechanism of LHNO [21].

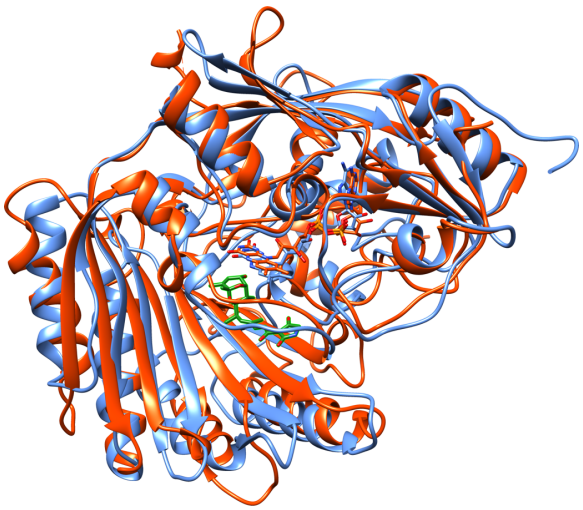
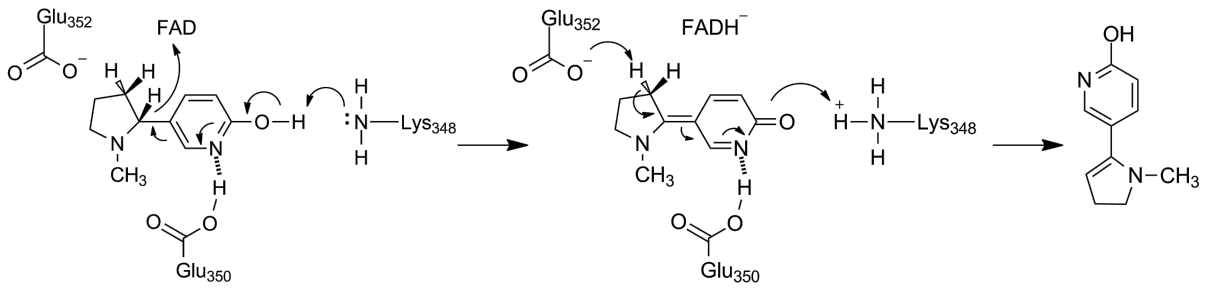
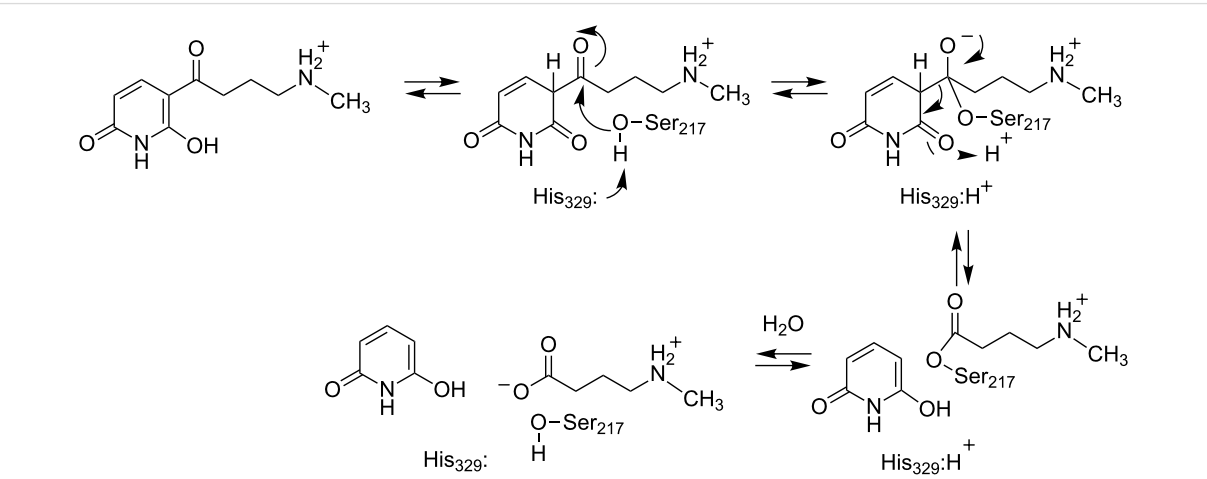


Figure 2: Overlay of the structures of DHNO (blue, pdb file 2bvf) and tirandamycin oxidase (orange, pdb file 2y3s), another member of the *p*-cresol methylhydroxylase/vanillyl oxidase family. The carbon atoms of tirandamycin are in green.

2,6-Dihydroxypseudooxynicotine hydrolase, the enzyme catalyzing the next step in the pyridine pathway, the cleavage of 2,6-dihydroxypseudooxynicotine to 2,6-dihydroxypyridine and *N*-methylaminobutyrate [34], was identified only after expression of a protein encoded by an open reading frame in pAO1 located next to the *kdhl* gene for the large subunit of 6-hydroxypseudooxynicotine dehydrogenase [35]. This protein was able to catalyze the cleavage of 2,6-dihydroxypseudooxynicotine without any added cofactors. A BLAST analysis of the sequence identified the protein as a member of the α/ß hydrolase family, which catalyzes a broad range of hydrolase and lyase reactions [36]. The subsequent determination of the crystal structure of the enzyme confirmed that it is an α/ß hydrolase, and mutagenesis identified the members of the catalytic triad as His_{329}, Ser_{217}, and Asp_{300} [37]. By analogy to other members of the family, the mechanism shown in Scheme 6 was proposed. There is an initial tautomerization to the diketo form of the substrate; Glu_{248} acts as both the initial proton acceptor and subsequent proton for this reaction. Nucleophilic attack of Ser_{217} on the substrate carbonyl followed by collapse of the tetrahedral intermediate generates an



Scheme 5: Proposed mechanism for DHNO [27].



Scheme 6: Mechanism of 2,6-dihydroxypseudooxynicotine hydrolase [37].

acyl–enzyme intermediate and the 2,6-dihydroxypyridine product. The subsequent hydrolysis of the acyl–enzyme intermediate then yields the *N*-methylaminobutyrate product. Other than the preliminary characterization of site-directed mutants of the protein, no mechanistic studies have been reported.

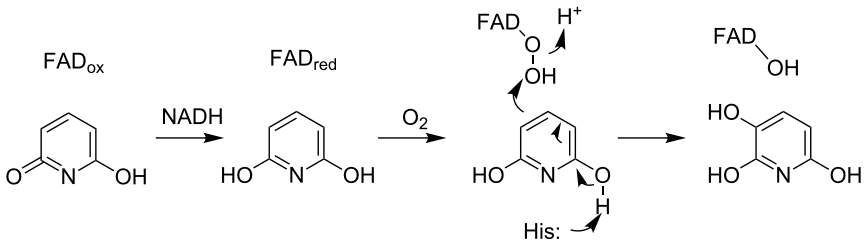
Two pathways have been identified for further metabolism of *N*-methylaminobutyrate. pAO1 contains separate genes for enzymes catalyzing the oxidative demethylation of *N*-methylaminobutyrate to form 4-aminobutyrate, *mabo*, and its oxidative deamination to form succinate semialdehyde and methylamine, *mao*. The expression of both proteins is regulated by nicotine [38,39], suggesting that both contribute in vivo. Mabo (γ-*N*-methylaminobutyrate demethylating oxidase) is similar in sequence to sarcosine dehydrogenase, and characterization of purified Mabo showed that it contains covalently-bound FAD and produces hydrogen peroxide as a product in addition to 4-aminobutryate [5,38]. Mabo also catalyzes the oxidative demethylation of sarcosine. Based on these results, the mechanism of the enzyme is similar to that of sarcosine oxidase, direct oxidation of the C–N bond of the substrate methyl group by hydride transfer [40]. The resulting 4-aminobutryrate is likely a substrate for a chromosomally-encoded aminotransferase, producing α-ketoglutarate and succinate semialdehyde. Mao (γ-*N*-methylaminobutyrate oxidase) contains noncovalently-bound flavin and catalyzes the oxidation of the other C–N bond of the methyl group in *N*-methylaminobutyrate to form methylamine and succinate semialdehyde, an MAO reaction [39]. While the k_{cat}/K_m value of Mao with 4-aminobutyrate is only 8% that of Mabo, *A. nicotinovorans* grown on [^{14}C]-nicotine produce [^{14}C]-methylamine, suggesting that Mao operates in vivo. Finally, pAO1 also contains the *sad* gene that codes for an NADP$^+$-dependent succinate semialdehyde dehydrogenase forming succinate as product [39].

2,3-Dihydroxypyridine 3-hydroxylase (2,6-DHPH), the enzyme converting 2,6-dihydroxypyridine to 2,3,6-trihydroxypyridine, has been cloned and characterized [10]. DHPH contains FAD and requires NADH and oxygen [41], and the sequence of the protein is similar to that of salicylate hydroxylase, although the sequence identity is only 21%. This allowed identification of the enzyme as a flavin-dependent phenol hydroxylase, a conclusion that was subsequently confirmed by the crystal structure of the enzyme (Figure 3) [42]. Based on the mechanism of this family of enzymes [43], the likely mechanism for this enzyme is as shown in Scheme 7. Flavin reduction by NADH is followed by the formation of the peroxyflavin hydroxylating intermediate. Attack of the substrate, activated by deprotonation of a substrate hydroxy group, on the peroxyflavin yields the hydroxylated product after a tautomerization. Two histidyl residues have been proposed to be involved in accepting the substrate proton. The details of further catabolism of 2,3,6-trihydroxypyridine are unclear. The compound can oxidatively dimerize to form nicotine blue [44], which is secreted into the medium. However, this has been proposed to be a byproduct, with the major pathway involving formation of maleamate, maleate, and fumarate [45].

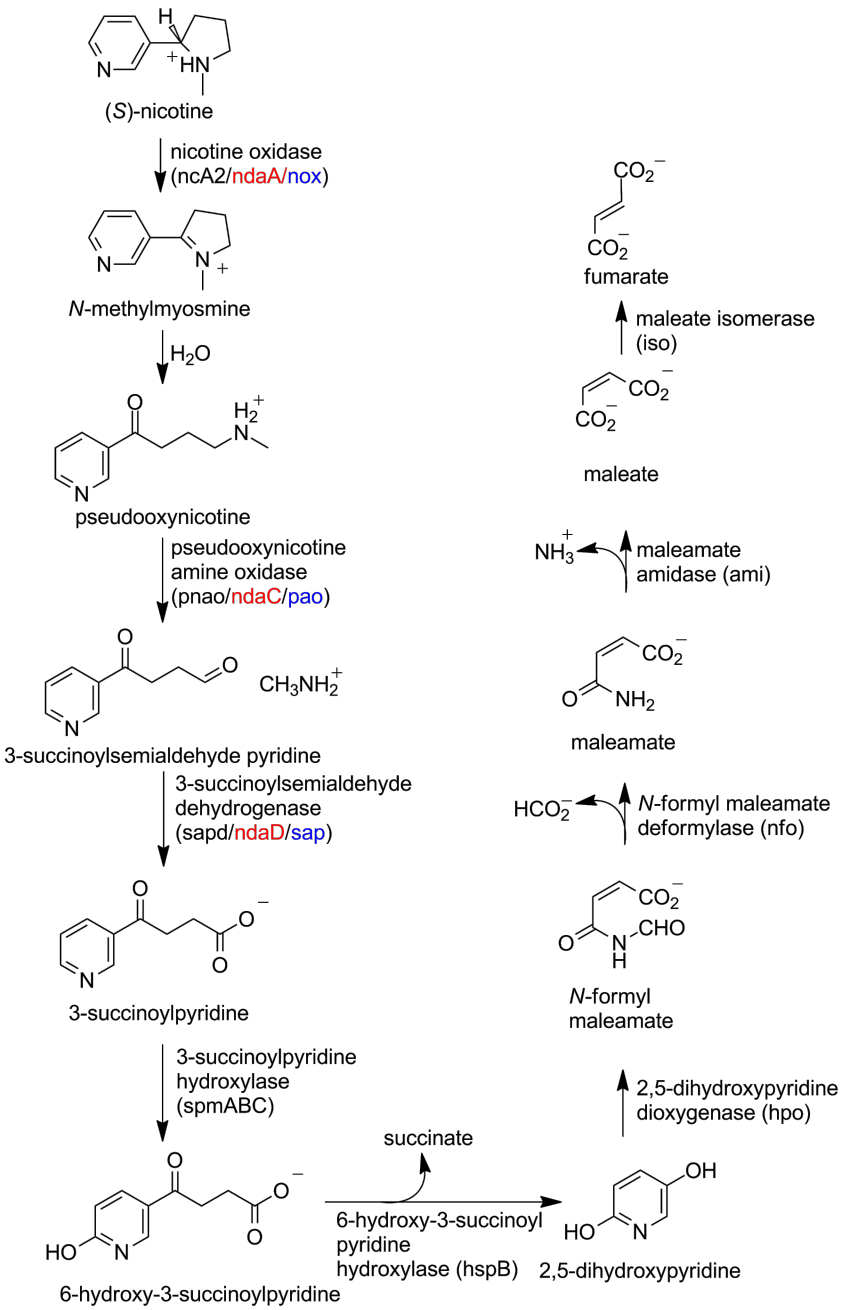
Figure 3: Overlay of structures of salicylate hydroxylase (orange, pdb file 5evy) and 2,3-dihydroxypyridine 3-hydroxylase (blue, pdb file 2vou). The salicylate bound to the latter is shown in green.

The pyrrolidine pathway

The metabolic pathway for nicotine degradation found in a number of pseudomonads (Scheme 8) [46-50] has been de-



Scheme 7: Mechanism of 2,3-dihydroxypyridine 3-hydroxylase [42].



Scheme 8: The pyrrolidine pathway for nicotine degradation by pseudomonads. The gene names for *P. putida* S16 (black), *P. putida* J5 (red), and *Pseudomonas* sp. HZN6 (blue) are in parentheses.

scribed as the pyrrolidine pathway. The initial oxidation of the pyrrolidine ring is catalyzed by the enzyme nicotine oxidase. In *P. putida* S16, the *nicA1* and *nicA2* genes produce separate enzymes that are both reported to have the ability to catalyze this reaction [48]. NicA1 was also reported to catalyze the subsequent oxidation of pseudooxynicotine to 3-succinylpyridine and methylamine, but no kinetic parameters for the two reactions were reported [51]. However, deletion of *nicA2* but not of *nicA1* prevents *P. putida* S16 from degrading nicotine, making it most likely that NicA2 is the relevant nicotine oxidase for this pathway [48]. In addition the amino acid sequence of NicA1 has no similarities to bacterial oxidases or dehydrogenases, instead resembling components of the bacterial electron transport chain. Thus, the function of NicA1 remains unclear, and NicA2 is likely the true nicotine oxidase. The *ndaA* gene in *P. putida* J5, required for degradation of nicotine by that organism, codes for

a protein that is 99% identical in sequence to that of NicA2 [52], so that NdaA is also likely to be a nicotine oxidase. The structure of NicA2 was recently determined, showing that the protein is a member of the MAO family with the same overall structure as LHNO (Figure 4) [53]. As is the case with LHNO, the NicA2-catalyzed reaction has generally been accepted to involve oxidation of a carbon–carbon bond in (*S*)-nicotine to form *N*-methylmyosmine. The recent evidence that the product of the oxidation of 6-hydroxynicotine by LHNO and DHNO arises from oxidation of a carbon–nitrogen bond [18,24] and the similarity of the active sites of LHNO and NicA2 to that of MAO makes it much more likely that the NicA2 instead catalyzes oxidation of the substrate carbon–nitrogen bond as shown in Scheme 4.

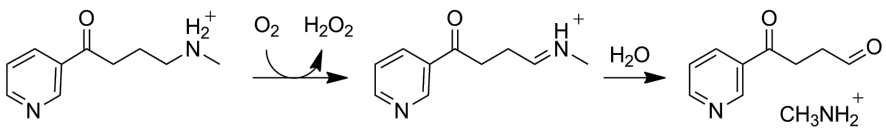
Figure 4: Overlay of the structure of LHNO (magenta, pdb file 3NG7) with that of NicA2 (magenta, pdb file 5ttj) B (green, pdb file 2FXU). The bound 6-hydroxynicotine is shown with green carbons.

Cloning and expression of the protein encoded by the *nox* gene of *Pseudomonas sp.* HZN6 showed that it also catalyzes oxidation of nicotine to pseudooxynicotine [54]. The sequence of the protein is most similar to that of LHNO and several members of the MAO family, consistent with Nox being a nicotine oxidase similar to NicA2. Nox is reported to be able to oxidize both stereoisomers of nicotine equally well, in contrast to the stereospecificity of LHNO and DHNO.

Pseudooxynicotine amine oxidase, the enzyme catalyzing the next step in the pathway, has been characterized from both *P. putida* S16 (Pnao) [48,55] and *P. putida* HZN6 (Pao) [47]. Both are FAD-containing enzymes whose sequences place them in the MAO family of flavoproteins. The sequence of NdaC from *P. putida* J5 is identical to that of Pnao, although the protein itself has not been characterized, and loss of *ndaC* eliminates the ability of cells to metabolize pseudooxynicotine [52]. In the case of Pnao the source of the oxygen in the 3-succinoylsemialdehyde pyridine product has been shown to be water [55], establishing the reaction catalyzed by the enzyme as shown in Scheme 9, with the hydrolytic step being nonenzymatic. This is essentially the same reaction as that catalyzed by *A. nicotinovorans* γ-*N*-methylaminobutyrate demethylating oxidase (Mabo).

E. coli expressing the *sap* gene from *Pseudomonas sp.* HZN6 will catalyze the $NADP^+$-dependent oxidation of 3-succinoylsemialdehyde pyridine to 3-succinoylpyridine [47], making SAP the likely 3-succinoylsemialdehyde dehydrogenase in the pyrrolidine pathway. The sequence of the enzyme identifies it as an aldehyde dehydrogenase [56], but the protein itself does not appear to have been characterized. In *P. putida* S16 and J5, sequence analyses have identified Spad and ndaD, respectively, as the likely 3-succinoylsemialdehyde pyridine dehydrogenases, and ndaD is required for *P. putida* J5 to convert 3-succinoylsemialdehyde pyridine to 3-succinoylpyridine [48,52].

Growth of *P. putida* S16 on nicotine results in increased expression of NicA2, Pnao, Sapd, SpmABC, and HspB, but not NicA1 or HspA [48]. The sequences of SpmA, SpmB, and SpmC are similar to those of nicotine dehydrogenase and other members of the xanthine dehydrogenase family. In addition, disrupting *spma* prevents *P. putida* S16 from converting 3-succinoylpyrimidine to 6-hydroxy-3-succinoylpyridine [48]. These results support the identification of SpmABC as a molybdopterin enzyme that catalyzes this step in the pathway. En-



Scheme 9: The pseudooxynicotine amine oxidase reaction.

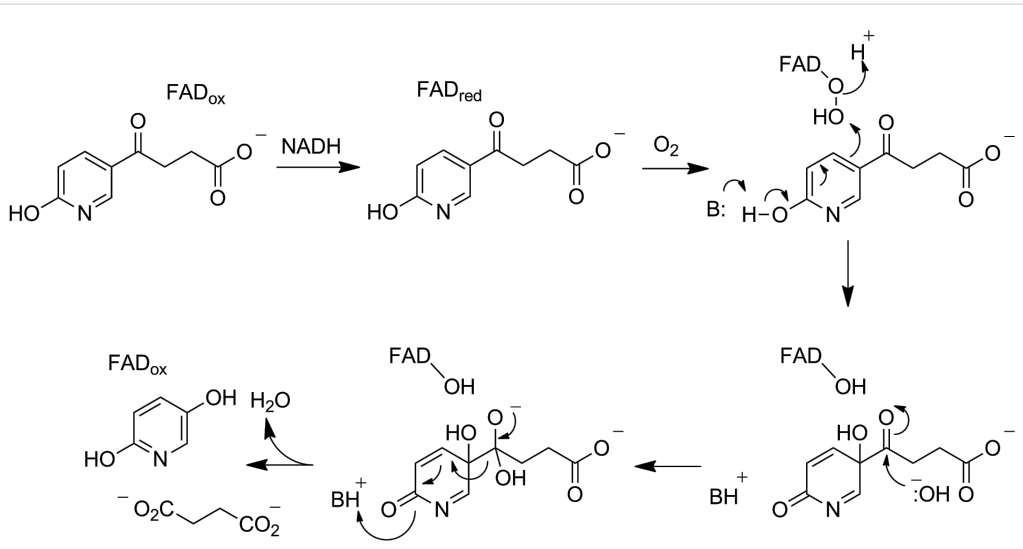
zymes with this activity do not appear to have been identified as yet for *P. putida* S5 and *Pseudomonas sp.* HZN6.

HspA in *P. putida* S16 was originally identified as a 6-hydroxy-3-succinoylpyridine hydroxylase catalyzing the formation of 2,5-dihydroxypyridine from 6-hydroxy-3-succinoylpyridine based on the location of the gene in a gene cluster that conferred on *E. coli* the ability to degrade nicotine to 2,5-dihydroxypyridine [57]. The sequence of the protein is not similar to that of any proteins with known functions. Purified recombinant HspA was reported to require NADH to catalyze the cleavage of 6-hydroxy-3-succinoylpyridine, but detailed kinetic analyses were not done. However, levels of HspA do not increase when *P. putida* S16 is grown on nicotine, while levels of HspB do [48]. Subsequent analysis of recombinant HspB showed that it contains FAD and catalyzes the NADH-dependent conversion of 6-hydroxy-3-succinoylpyridine to 2,5-dihydroxypyridine [58,59]. The new oxygen atom in 2,5-dihydroxypyridine comes from O_2, while that in succinate comes from H_2O. The sequence of HspB is closest to those of a number of FAD-dependent hydroxylases, and a peroxyflavin was detected in stopped-flow analyses of the enzyme-catalyzed reaction. These results led to the mechanism shown in Scheme 10 for HspB. A similar enzyme has been isolated from *Pseudomonas* sp. ZZ-5 [60].

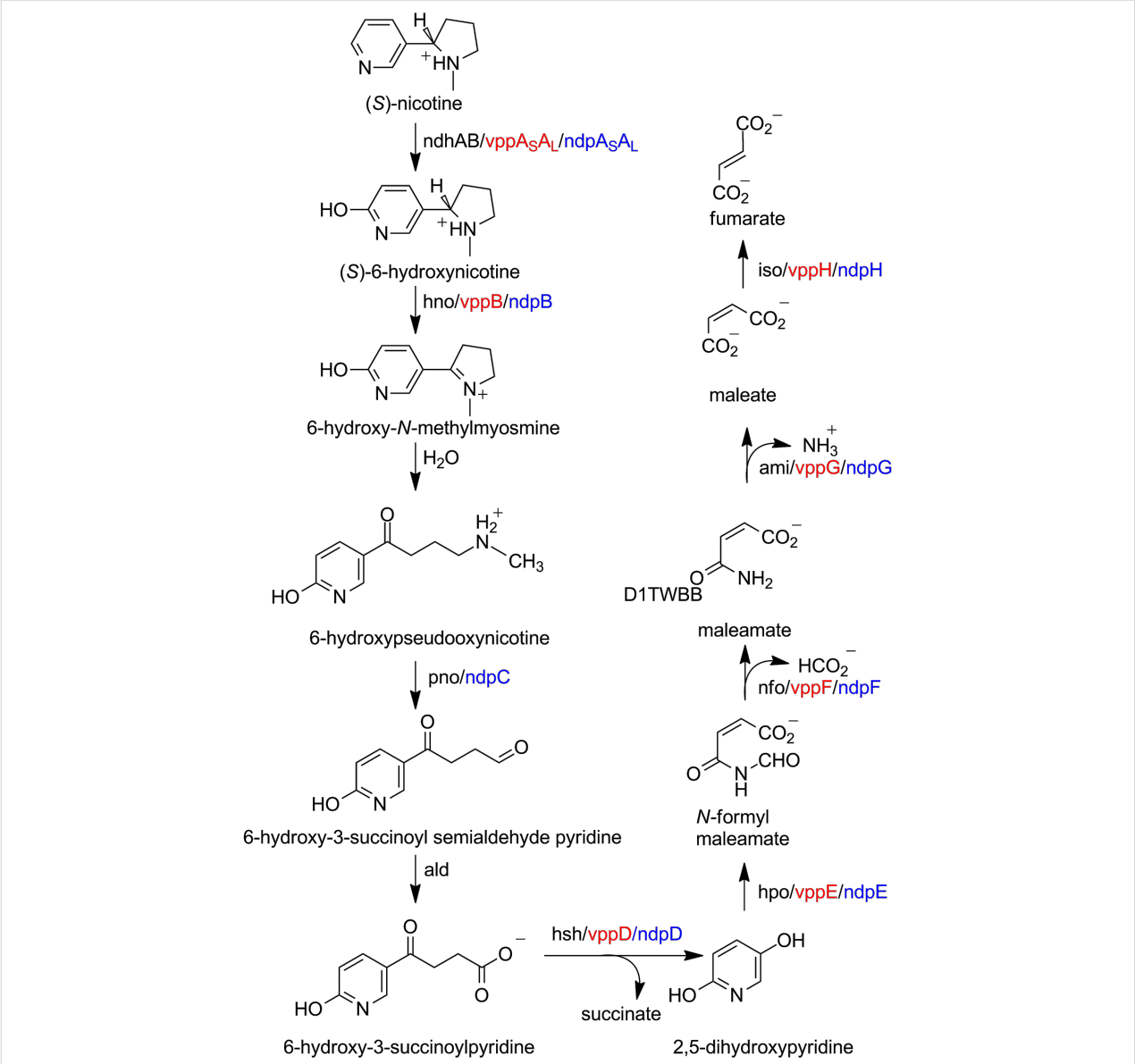
The subsequent steps in metabolism of 2,5-dihydropyridine by *P. putida* S16 were identified when the gene cluster *nic2* that contained *hspb* was sequenced, with the demonstration that incorporation of *nic2* into *E. coli* allowed cells to convert 6-hydroxy-3-succinoylpyridine to fumarate [61]. In addition to hspB and an unidentified open reading frame, four genes could be identified by sequence analyses as likely to code for proteins catalyzing the final steps in nicotine catabolism. These four proteins were expressed individually in *E. coli* and characterized. Hpo catalyzes the Fe(II)-dependent formation of *N*-formylmaleamate from 2,5-dihydroxypyridine in the absence of other cofactors or substrates; it was designated DHP dioxygenase. Both oxygen atoms in the product come from O_2 and mutagenesis of the predicted iron ligand His257, His310, or Asp312 results in loss of activity, consistent with Hpo being a non-heme Fe(II)-dependent dioxygenase [62]. Nfo catalyzes the formation of maleamate from *N*-formylmaleamate and was designated *N*-formylmaleamate deformylase; its sequence identifies it as a member of the α/ß hydrolase superfamily [36]. Ami is a maleamate amidase that catalyzes the hydrolysis of maleamate to maleic acid plus ammonium; it also belongs to the α/ß hydrolase superfamily. Finally, Iso catalyzes the reversible isomerization of maleate to fumarate. Orthologues of all four of these enzymes have been identified as being involved in the metabolism of nicotinic acid by *P. putida* KT25440, which begins with the hydroxylation of nicotinic acid by the molybdopterin enzyme NicAB to form 6-hydroxynicotinic acid and its subsequent conversion to 2,5-dihydroxypyridine by the NADH- and FAD-dependent hydroxylase NicC [63,64].

The hybrid pathway

While the pyridine and pyrrolidine pathways are the best understand reactions by which bacteria degrade nicotine, additional pathways continue to be discovered. The best-characterized is a hybrid of the pyridine and pyrrolidine pathways (Scheme 11). Based on phylogenetic analysis, the pathway is more closely related to the pyrrolidine pathway, with both found predominantly in Gram-negative bacteria [65]. This pathway is best characterized for *Agrobacter tumefaciens* S33, *Ochrobactrum*



Scheme 10: Mechanism of HspB [59].



Scheme 11: Hybrid pyridine/pyrrolidine pathway for nicotine metabolism in *Agrobacter tumefaciens* S33 (black), *Ochrobactrum* sp. SJY1 (red), and *Sphingomonas melonis* Ty (blue).

sp. SJY1, *Sphingomonas melonis* TY, and *Shinella* sp. HZN7, but has been identified in other bacteria as well [65-70]. The pathway begins with the hydroxylation of nicotine, as in the pyridine pathway, but diverges after the formation of 6-hydroxypseudooxynicotine. The oxidative deamination of 6-hydroxypseudooxynicotine yields 6-hydroxy-3-succinoylsemialdehyde pyridine, an intermediate that is not present in the other two pathways; its oxidation forms 6-hydroxy-3-succinoylpyridine, which is processed further as in the pyrrolidine pathway. Elucidation of the hybrid pathway has relied on identification of intermediates and on comparison of gene sequences with those coding for enzymes known to be involved in nicotine catabolism in the pyridine and pyrrolidine pathways.

The identification of 6-hydroxy-L-nicotine, 6-hydroxy-*N*-methylmyosmine, 6-hydroxypseudooxynicotine, 6-hydroxy-3-succinoylpyridine, and 2,5-dihydroxypyridine as metabolites in cells of *A. tumefaciens* S33 degrading nicotine provided the initial evidence for a pathway different from those shown in Scheme 1 and Scheme 8 [66]. The complete genome of *A. tumefaciens* S33 was recently sequenced, allowing identification of candidate genes for all of the steps for the hybrid pathway for nicotine degradation in that organism (Scheme 11) [69]. The nicotine dehydrogenase Ndh and the 6-hydroxypseudooxynicotine oxidase Pno, one of the two novel enzymes in this pathway, have both been purified and characterized; they are reported to form a complex [71]. The sequences of the two en-

zymes identify Ndo as a member of the family of molybdopterin enzymes such as xanthine dehydrogenase and Pno as a member of the trimethylamine dehydrogenase family of flavoproteins [72]. Consistent with this identification, purified Pno contains FMN and a 4Fe/4S center. Preliminary kinetics have been reported for both enzymes. 6-LHNO activity has been detected in crude cell lysates of *A. tumefaciens* S33 grown on nicotine, but the pure enzyme has not been described [66]. The NADH-dependent 6-hydroxy-3-succinoylpyridine hydroxylase Hsh has been partially purified from this organism; this enzyme is likely an FAD-dependent hydroxylase similar to HspB, but succinate has not been shown to be a product in this case [66].

6-Hydroxy-L-nicotine, 6-hydroxy-*N*-methylmyosmine, 6-hydroxypseudooxynicotine, 6-hydroxy-3-succinoylpyridine, and 2,5-dihydroxypyridine have also been isolated from cells of *Ochrobactrum* sp. SJY1 growing on nicotine [68]. The sequencing of its genome allowed identification of several of the genes involved in nicotine degradation (Scheme 11) [73]. VppB, VppD, and VppE have all been expressed in recombinant form [68]. VppB is a flavin amine oxidase that catalyzes the oxidation of 6-hydroxynicotine, establishing it as an LHNO, although the sequence of the protein is closer to that of *P. putida* S16 NicA2 than LHNO from *A. nicotinovorans*. VppD is an NAD(P)-dependent flavin monooxygenase whose sequence is 62% identical to that of *P. putida* S16 HspB. The crude recombinant VppE catalyzes the iron-dependent oxidation of 2,5-dihydroxypyridine to *N*-formylmaleamate, the same reaction as is catalyzed by the dioxygenase Hpo from *P. putida* S16.

In *Sphingomonas melonis* TY, the genes for the metabolism of nicotine are found in the *ndp* gene cluster [65]. The mRNA levels for *ndpA-H* all increase 10 to 100-fold upon growth in the presence of nicotine. Sequence similarities of >35% in all cases to genes in *P. putida* S16 involved in nicotine metabolism suggested that the roles of each are as those shown in Scheme 11. NdpA-D were all expressed in recombinant form and shown to confer on cells the ability to catalyze the proposed reactions, confirming the identification of NdpA as a nicotine dehydrogenase, NdpB as an LHNO, NdpC as an oxidative demethylase, and NdpD as 6-hydroxy-3-succinoylpyridine 3-monooxygenase. The enzyme catalyzing formation of 6-hydroxy-3-succinoylpyridine from the aldehyde was not identified in the *ndp* cluster; this activity was attributed to a non-specific semialdehyde dehydrogenase.

Characterization of the enzymes involved in the hybrid pathway in *Shinella* sp. HZN7 is less complete. 6-Hydroxynicotine, 6-hydroxy-*N*-methylmyosmine, 6-hydroxypseudooxynicotine, 6-hydroxy-3-succinoylpyridine, and 2,5-dihydroxypyridine have been confirmed as intermediates in the degradation of nicotine by this organism [67]. This bacterium is also able to utilize 2,5-dihydroxypyridine as a sole carbon source, establishing the presence of the complete pathway. 6-Hydroxy-3-succinoyl semialdehyde pyridine was not reported as a detectable intermediate, but it might not have accumulated to sufficient levels for detection. The genes *nctA1* and *nctA2* code for proteins with sequences identical to the nicotine hydroxylase VppAB from *Ochrobactrum* sp. SJY1 [74]. The gene *nctB* was identified as required for nicotine degradation using genetic approaches; NctB was expressed in *E. coli* and the purified protein shown to be an LHNO similar to the enzyme from *A. nicotinovorans* in its kinetic properties [18,75]. The genes responsible for the other enzymes were tentatively identified by comparison with the sequences of the enzymes from *Ochrobactrum* sp. SJY1 and *A. tumefaciens* when the complete genome of *Shinella* sp. HZN7 was sequenced [76].

Conclusion

This review has attempted to summarize our present understanding of the microbial metabolism of nicotine, with an emphasis on the enzymes involved. It has not attempted to address the less understood fungal metabolism of nicotine. Elucidation of the details of nicotine metabolism remains one of intense investigation, and the rapid increase in genomic sequences means that additional organisms capable of degrading nicotine are frequently being described. Many of the enzymes involved are poorly characterized even if mechanisms can be proposed for them based on their homology to known families of enzymes, and not all of the enzymes have been identified in some cases. Still, these enzymes are already being used to produce new synthetically catalysts, while the pathways are being retooled to produce useful synthetic intermediates.

Acknowledgements

Research from the author's laboratory described here was supported in part by the NIH (GM58698) and The Welch Foundation (AQ-1245).

ORCID® iDs

Paul F. Fitzpatrick - https://orcid.org/0000-0002-8486-2024

References

1. Liu, J.; Ma, G.; Chen, T.; Hou, Y.; Yang, S.; Zhang, K.-Q.; Yang, J. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 3775–3785. doi:10.1007/s00253-015-6525-1
2. Yu, W.; Wang, R.; Li, H.; Liang, J.; Wang, Y.; Huang, H.; Xie, H.; Wang, S. *Biotechnol. Biofuels* **2017,** *10,* No. 288. doi:10.1186/s13068-017-0976-9
3. Brandsch, R. *Appl. Microbiol. Biotechnol.* **2006,** *69,* 493–498. doi:10.1007/s00253-005-0226-0

4. Brandsch, R.; Hinkkanen, A. E.; Decker, K. *Arch. Microbiol.* **1982,** *132,* 26–30. doi:10.1007/BF00690812
5. Igloi, G. L.; Brandsch, R. *J. Bacteriol.* **2003,** *185,* 1976–1986. doi:10.1128/JB.185.6.1976-1986.2003
6. Ganas, P.; Sachelaru, P.; Mihasan, M.; Igloi, G. L.; Brandsch, R. *Arch. Microbiol.* **2008,** *189,* 511–517. doi:10.1007/s00203-007-0340-8
7. Hochstein, L. I.; Rittenberg, S. C. *J. Biol. Chem.* **1959,** *234,* 156–160.
8. Hille, R.; Hall, J.; Basu, P. *Chem. Rev.* **2014,** *114,* 3963–4038. doi:10.1021/cr400443z
9. Grether-Beck, S.; Igloi, G. L.; Pust, S.; Schilz, E.; Decker, K.; Brandsch, R. *Mol. Microbiol.* **1994,** *13,* 929–936. doi:10.1111/j.1365-2958.1994.tb00484.x
10. Baitsch, D.; Sandu, C.; Brandsch, R.; Igloi, G. L. *J. Bacteriol.* **2001,** *183,* 5262–5267. doi:10.1128/JB.183.18.5262-5267.2001
11. Menéndez, C.; Otto, A.; Igloi, G.; Nick, P.; Brandsch, R.; Schubach, B.; Böttcher, B.; Brandsch, R. *Eur. J. Biochem.* **1997,** *250,* 524–531. doi:10.1111/j.1432-1033.1997.0524a.x
12. Gries, F. A.; Decker, K.; Brühmueller, M. *Hoppe-Seyler's Z. Physiol. Chem.* **1961,** *325,* 229–241. doi:10.1515/bchm2.1961.325.1.229
13. Dang Dai, V.; Decker, K.; Sund, H. *Eur. J. Biochem.* **1968,** *4,* 95–102. doi:10.1111/j.1432-1033.1968.tb00177.x
14. Schenk, S.; Hoelz, A.; Krauß, B.; Decker, K. *J. Mol. Biol.* **1998,** *284,* 1323–1339. doi:10.1006/jmbi.1998.2227
15. Kachalova, G.; Decker, K.; Holt, A.; Bartunik, H. D. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 4800–4805. doi:10.1073/pnas.1016684108
16. Gaweska, H.; Fitzpatrick, P. F. *Biomol. Concepts* **2011,** *2,* 365–377. doi:10.1515/BMC.2011.030
17. Decker, K.; Dai, V. D. *Eur. J. Biochem.* **1967,** *3,* 132–138. doi:10.1111/j.1432-1033.1967.tb19507.x
18. Fitzpatrick, P. F.; Chadegani, F.; Zhang, S.; Roberts, K. M.; Hinck, C. S. *Biochemistry* **2016,** *55,* 697–703. doi:10.1021/acs.biochem.5b01325
19. Fitzpatrick, P. F.; Chadegani, F.; Zhang, S.; Dougherty, V. *Biochemistry* **2017,** *56,* 869–875. doi:10.1021/acs.biochem.6b01160
20. Fitzpatrick, P. F. *Arch. Biochem. Biophys.* **2010,** *493,* 13–25. doi:10.1016/j.abb.2009.07.019
21. Kachalova, G. S.; Bourenkov, G. P.; Mengesdorf, T.; Schenk, S.; Maun, H. R.; Burghammer, M.; Riekel, C.; Decker, K.; Bartunik, H. D. *J. Mol. Biol.* **2010,** *396,* 785–799. doi:10.1016/j.jmb.2009.12.009
22. Armstrong, D. W.; Wang, X.; Ercal, N. *Chirality* **1998,** *10,* 587–591. doi:10.1002/(SICI)1520-636X(1998)10:7<587::AID-CHIR6>3.0.CO;2-#
23. Decker, K.; Eberwein, H.; Gries, F. A.; Bruehmueller, M. *Biochem. Z.* **1961,** *334,* 227–244.
24. Heath, R. S.; Pontini, M.; Bechi, B.; Turner, N. J. *ChemCatChem* **2014,** *6,* 996–1002. doi:10.1002/cctc.201301008
25. Fraaije, M. W.; van Berkel, W. J. H.; Benen, J. A. E.; Visser, J.; Mattevi, A. *Trends Biochem. Sci.* **1998,** *23,* 206–207. doi:10.1016/S0968-0004(98)01210-9
26. Möhler, H.; Brühmüller, M.; Decker, K. *Eur. J. Biochem.* **1972,** *29,* 152–155. doi:10.1111/j.1432-1033.1972.tb01969.x
27. Koetter, J. W. A.; Schulz, G. E. *J. Mol. Biol.* **2005,** *352,* 418–428. doi:10.1016/j.jmb.2005.07.041
28. Fraaije, M. W.; van Berkel, W. J. H. *J. Biol. Chem.* **1997,** *272,* 18111–18116. doi:10.1074/jbc.272.29.18111
29. Fraaije, M. W.; Veeger, C.; Van Berkel, W. J. H. *Eur. J. Biochem.* **1995,** *234,* 271–277. doi:10.1111/j.1432-1033.1995.271_c.x
30. Ewing, T. A.; Fraaije, M. W.; Mattevi, A.; van Berkel, W. J. H. *Arch. Biochem. Biophys.* **2017,** *632,* 104–117. doi:10.1016/j.abb.2017.06.022
31. Holm, L.; Rosenström, P. *Nucleic Acids Res.* **2010,** *38,* W545–W549. doi:10.1093/nar/gkq366
32. Richardson, S. H.; Rittenberg, S. C. *J. Biol. Chem.* **1961,** *236,* 964–967.
33. Sachelaru, P.; Schiltz, E.; Brandsch, R. *Appl. Environ. Microbiol.* **2006,** *72,* 5126–5131. doi:10.1128/AEM.00437-06
34. Gherna, R. L.; Richardson, S. H.; Rittenberg, S. C. *J. Biol. Chem.* **1965,** *240,* 3669–3674.
35. Sachelaru, P.; Schiltz, E.; Igloi, G. L.; Brandsch, R. *J. Bacteriol.* **2005,** *187,* 8516–8519. doi:10.1128/JB.187.24.8516-8519.2005
36. Rauwerdink, A.; Kazlauskas, R. J. *ACS Catal.* **2015,** *5,* 6153–6176. doi:10.1021/acscatal.5b01539
37. Schleberger, C.; Sachelaru, P.; Brandsch, R.; Schulz, G. E. *J. Mol. Biol.* **2007,** *367,* 409–418. doi:10.1016/j.jmb.2006.12.068
38. Chiribau, C. B.; Sandu, C.; Fraaije, M.; Schiltz, E.; Brandsch, R. *Eur. J. Biochem.* **2004,** *271,* 4677–4684. doi:10.1111/j.1432-1033.2004.04432.x
39. Chiribau, C.-B.; Mihasan, M.; Ganas, P.; Igloi, G. L.; Artenie, V.; Brandsch, R. *FEBS J.* **2006,** *273,* 1528–1536. doi:10.1111/j.1742-4658.2006.05173.x
40. Ralph, E. C.; Hirschi, J. S.; Anderson, M. A.; Cleland, W. W.; Singleton, D. A.; Fitzpatrick, P. F. *Biochemistry* **2007,** *46,* 7655–7664. doi:10.1021/bi700482h
41. Holmes, P. E.; Rittenberg, S. C. *J. Biol. Chem.* **1972,** *247,* 7622–7627.
42. Treiber, N.; Schulz, G. E. *J. Mol. Biol.* **2008,** *379,* 94–104. doi:10.1016/j.jmb.2008.03.032
43. Huijbers, M. M.; Montersino, S.; Westphal, A. H.; Tischler, D.; van Berkel, W. J. H. *Arch. Biochem. Biophys.* **2014,** *544,* 2–17. doi:10.1016/j.abb.2013.12.005
44. Knackmuss, H.-J.; Beckmann, W. *Arch. Mikrobiol.* **1973,** *90,* 167–169. doi:10.1007/BF00414521
45. Kaiser, J. P.; Feng, Y.; Bollag, J. M. *Microbiol. Rev.* **1996,** *60,* 483–498.
46. Wang, S. N.; Liu, Z.; Tang, H. Z.; Meng, J.; Xu, P. *Microbiology (London, U. K.)* **2007,** *153,* 1556–1565. doi:10.1099/mic.0.2006/005223-0
47. Qiu, J.; Ma, Y.; Wen, Y.; Chen, L.; Wu, L.; Liu, W. *Appl. Environ. Microbiol.* **2012,** *78,* 2154–2160. doi:10.1128/AEM.07025-11
48. Tang, H.; Wang, L.; Wang, W.; Yu, H.; Zhang, K.; Yao, Y.; Xu, P. *PLoS Genet.* **2013,** *9,* e1003923. doi:10.1371/journal.pgen.1003923
49. Raman, G.; Sakthivel, N.; Park, S. *Genome Announce.* **2015,** *3,* No. e01162-14. doi:10.1128/genomeA.01162-14
50. Li, A.; Qiu, J.; Chen, D.; Ye, J.; Wang, Y.; Tong, L.; Jiang, J.; Chen, J. *Mar. Drugs* **2017,** *15,* No. 156. doi:10.3390/md15060156
51. Tang, H.; Wang, L.; Meng, X.; Ma, L.; Wang, S.; He, X.; Wu, G.; Xu, P. *Appl. Environ. Microbiol.* **2009,** *75,* 772–778. doi:10.1128/AEM.02300-08
52. Xia, Z.; Zhang, W.; Lei, L.; Liu, X.; Wei, H.-L. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 6503–6514. doi:10.1007/s00253-015-6529-x
53. Tararina, M. A.; Janda, K. D.; Allen, K. N. *Biochemistry* **2016,** *55,* 6595–6598. doi:10.1021/acs.biochem.6b00963
54. Qiu, J.; Ma, Y.; Zhang, J.; Wen, Y.; Liu, W. *Appl. Environ. Microbiol.* **2013,** *79,* 2164–2171. doi:10.1128/AEM.03824-12
55. Hu, H.; Wang, W.; Tang, H.; Xu, P. *Sci. Rep.* **2015,** *5,* No. 17770. doi:10.1038/srep17770

56. Liu, Z.-J.; Sun, Y.-J.; Rose, J.; Chung, Y.-J.; Hsiao, C.-D.; Chang, W.-R.; Kuo, I.; Perozich, J.; Lindahl, R.; Hempel, J.; Wang, B.-C. *Nat. Struct. Biol.* **1997,** *4,* 317–326. doi:10.1038/nsb0497-317
57. Tang, H.; Wang, S.; Ma, L.; Meng, X.; Deng, Z.; Zhang, D.; Ma, C.; Xu, P. *Appl. Environ. Microbiol.* **2008,** *74,* 1567–1574. doi:10.1128/AEM.02529-07
58. Tang, H.; Yao, Y.; Zhang, D.; Meng, X.; Wang, L.; Yu, H.; Ma, L.; Xu, P. *J. Biol. Chem.* **2011,** *286,* 39179–39187. doi:10.1074/jbc.M111.283929
59. Yu, H.; Hausinger, R. P.; Tang, H.-Z.; Xu, P. *J. Biol. Chem.* **2014,** *289,* 29158–29170. doi:10.1074/jbc.M114.558049
60. Wei, T.; Zang, J.; Zheng, Y.; Tang, H.; Huang, S.; Mao, D. *Catalysts* **2017,** *7,* No. 257. doi:10.3390/catal7090257
61. Tang, H.; Yao, Y.; Wang, L.; Yu, H.; Ren, Y.; Wu, G.; Xu, P. *Sci. Rep.* **2012,** *2,* No. 377. doi:10.1038/srep00377
62. Kovaleva, E. G.; Lipscomb, J. D. *Nat. Chem. Biol.* **2008,** *4,* 186–193. doi:10.1038/nchembio.71
63. Jiménez, J. I.; Canales, Á.; Jiménez-Barbero, J.; Ginalski, K.; Rychlewski, L.; García, J. L.; Díaz, E. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 11329–11334. doi:10.1073/pnas.0802273105
64. Belda, E.; van Heck, R. G. A.; Lopez-Sanchez, M. J.; Cruveiller, S.; Barbe, V.; Fraser, C.; Klenk, H.-P.; Petersen, J.; Morgat, A.; Nikel, P. I.; Vallenet, D.; Rouy, Z.; Sekowska, A.; Martins Dos Santos, V. A. P.; de Lorenzo, V.; Danchin, A.; Médigue, C. *Environ. Microbiol.* **2016,** *18,* 3403–3424. doi:10.1111/1462-2920.13230
65. Wang, H.; Zhi, X.-Y.; Qiu, J.; Shi, L.; Lu, Z. *Front. Microbiol.* **2017,** *8,* No. 337. doi:10.3389/fmicb.2017.00337
66. Wang, S.; Huang, H.; Xie, K.; Xu, P. *Appl. Microbiol. Biotechnol.* **2012,** *95,* 1567–1578. doi:10.1007/s00253-012-4007-2
67. Ma, Y.; Wei, Y.; Qiu, J.; Wen, R.; Hong, J.; Liu, W. *Appl. Microbiol. Biotechnol.* **2014,** *98,* 2625–2636. doi:10.1007/s00253-013-5207-0
68. Yu, H.; Tang, H.; Zhu, X.; Li, Y.; Xu, P. *Appl. Environ. Microbiol.* **2015,** *81,* 272–281. doi:10.1128/AEM.02265-14
69. Huang, H.; Yu, W.; Wang, R.; Li, H.; Xie, H.; Wang, S. *Sci. Rep.* **2017,** *7,* No. 4813. doi:10.1038/s41598-017-05320-1
70. Li, J.; Qian, S.; Xiong, L.; Zhu, C.; Shu, M.; Wang, J.; Jiao, Y.; He, H.; Zhang, F.; Linhardt, R. J.; Zhong, W. *Front. Microbiol.* **2017,** *8,* No. 2085. doi:10.3389/fmicb.2017.02085
71. Li, H.; Xie, K.; Yu, W.; Hu, L.; Huang, H.; Xie, H.; Wang, S. *Appl. Environ. Microbiol.* **2016,** *82,* 1745–1755. doi:10.1128/AEM.03909-15
72. Fitzpatrick, P. F. Amine and amino acid oxidases and dehydrogenases. In *Handbook of Flavoproteins;* Miller, S.; Hille, R.; Palfey, B. A., Eds.; Walter de Gruyter: Berlin, 2013; pp 119–138.
73. Yu, H.; Li, Y.; Tang, H.; Xu, P. *Genome Announce*. **2014,** *2,* No. e00720-14. doi:10.1128/genomeA.00720-14
74. Qiu, J.; Li, N.; Lu, Z.; Yang, Y.; Ma, Y.; Niu, L.; He, J.; Liu, W. *Appl. Microbiol. Biotechnol.* **2016,** *100,* 10019–10029. doi:10.1007/s00253-016-7805-0
75. Qiu, J.; Wei, Y.; Ma, Y.; Wen, R.; Wen, Y.; Liu, W. *Appl. Environ. Microbiol.* **2014,** *80,* 5552–5560. doi:10.1128/AEM.01312-14
76. Qiu, J.; Yang, Y.; Zhang, J.; Wang, H.; Ma, Y.; He, J.; Lu, Z. *Front. Microbiol.* **2016,** *7,* No. 1348. doi:10.3389/fmicb.2016.01348

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Thermophilic phosphoribosyltransferases *Thermus thermophilus* HB27 in nucleotide synthesis

Ilja V. Fateev, Ekaterina V. Sinitsina, Aiguzel U. Bikanasova, Maria A. Kostromina, Elena S. Tuzova, Larisa V. Esipova, Tatiana I. Muravyova, Alexei L. Kayushin, Irina D. Konstantinova* and Roman S. Esipov*

Full Research Paper

Open Access

Address:
Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Miklukho-Maklaya Str., 16/10, Moscow, GSP-7, 117997, Russia

Email:
Irina D. Konstantinova* - kid1968@yandex.ru; Roman S. Esipov* - esipov@ibch.ru

* Corresponding author

Keywords:
adenine phosphoribosyltransferase; catalysis; enzyme; hypoxanthine phosphoribosyltransferase; multi-enzyme cascade; nucleotides; thermophiles

Beilstein J. Org. Chem. **2018,** *14,* 3098–3105.
doi:10.3762/bjoc.14.289

Received: 05 July 2018
Accepted: 27 November 2018
Published: 21 December 2018

This article is part of the thematic issue "Enzymes in chemical transformations".

Guest Editor: K. N. Allen

Abstract

Phosphoribosyltransferases are the tools that allow the synthesis of nucleotide analogues using multi-enzymatic cascades. The recombinant adenine phosphoribosyltransferase (*Tth*APRT) and hypoxanthine phosphoribosyltransferase (*Tth*HPRT) from *Thermus thermophilus* HB27 were expressed in *E.coli* strains and purified by chromatographic methods with yields of 10–13 mg per liter of culture. The activity dependence of *Tth*APRT and *Tth*HPRT on different factors was investigated along with the substrate specificity towards different heterocyclic bases. The kinetic parameters for *Tth*HPRT with natural substrates were determined. Two nucleotides were synthesized: 9-(β-D-ribofuranosyl)-2-chloroadenine 5'-monophosphate (2-Cl-AMP) using *Tth*APRT and 1-(β-D-ribofuranosyl)pyrazolo[3,4-*d*]pyrimidine-4-one 5'-monophosphate (Allop-MP) using *Tth*HPRT.

Introduction

Bacterial phosphoribosyltransferases are used in multi-enzymatic cascades that perform nucleotide synthesis de novo [1,2]. Recently, we reported on the possibility of cascade synthesis, where enzymes of thermophilic microorganisms *Thermus thermophilus* HB27 (phosphoribosylpyrophosphate synthetase – PRPPS and adenine phosphoribosyltransferase – APRT) and *Thermus* sp. 2.9 (ribokinase – RK) carry out successive transformations of ribose and adenine heterocyclic bases into the corresponding nucleotides (Figure 1). The use of thermophilic phosphoribosyltransferases allows carrying out reactions at a higher temperature, so the concentrations of heterocyclic bases can be increased [1-3].

There is great interest in the development of multi-enzymatic cascades [4-9] for the preparation of nucleosides and nucleotides due to the regio- and stereospecificity of enzymes

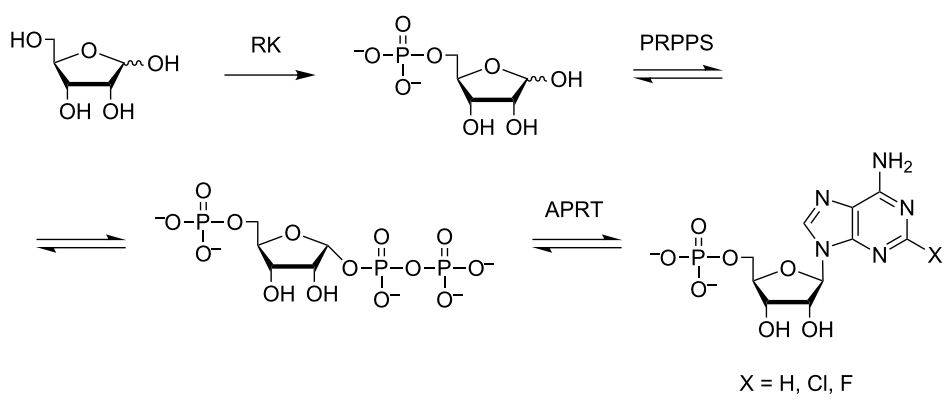


Figure 1: A multi-enzymatic synthesis of modified adenosine -5'-monophosphates.

[4,10,11], performing metabolic transformations of substrates. Phosphoribosyltransferases are increasingly being widely used as key enzymes in multi-enzymatic systems [2]. The substrate specificity of APRT limits the number of possible nucleotides that can be synthesized. Thus, for *Thermus thermophilus* HB27 APRT (*Tth*APRT), nucleotide synthesis is limited to the closest structural homologs of adenine (Table 1) [1].

Table 1: Substrates for *Tth*APRT.

Base	Conversion into nucleotide (24 h, %)[a]
2,6-diaminopurine	16.80
2-chloroadenine	97.58
2-fluoroadenine	36.50
adenine	50.02
2-methoxyadenine	60.88
N1-methyladenine	78.2
N6-benzyladenine	1.87
2-aminobenzimidazole	0.09

[a]Reaction mixtures (0.5 mL, 20 mM Tris-HCl, pH 8.0, 75 °C) contained 0.4 mM heterocyclic base, 0.4 mM PRPP, 0–5 mM $MgCl_2$, 1.25 μg *Tth*APRT [1].

Unfortunately, 1,2,4-triazole-3-carboxamide, its analogues, guanine, hypoxanthine, and 7-deazapurins are not substrates for *Tth*APRT. This severely limits the usability of multi-enzymatic cascades in the synthesis of nucleotides, including the modified ones.

To expand the possible repertoire of nucleotides that could be synthesized, we obtained the recombinant form of hypoxathine phosphoribosyltransferase *Thermus thermophilus* (*Tth*HPRT), investigated its substrate specificity and optimal conditions for catalytic activity, and determined the kinetic parameters of the enzyme. A comparative study of the substrate specificity of *Tth*APRT and *Tth*HPRT was performed to determine the usability of thermophilic transferases in nucleotide synthesis. A scheme of purine nucleotide synthesis using *Tth*APRT and *Tth*HPRT is shown in Figure 2.

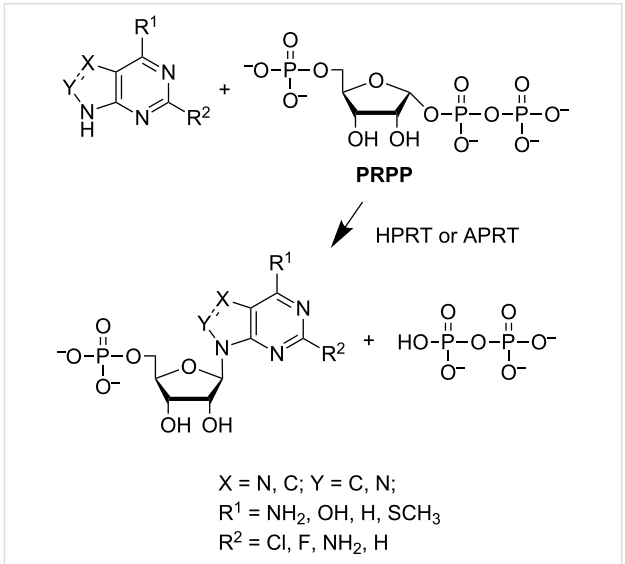


Figure 2: Nucleotide synthesis using phosphoribosyltransferases.

Results and Discussion

Genes TT_RS08985 and TT_RS06315 from *T. thermophilus* HB27, coding *Tth*HPRT and *Tth*APRT, were cloned into expression plasmid vectors pET 23a+ and pET 23d+, respectively. The resulting recombinant plasmid pER-*Tth*HPRT contained fusion gene HPRT-HisTag coding *Tth*HPRT with a C-terminal His-Tag. The resulting recombinant plasmid pER-*Tth*APRT contained the gene APRT coding *Tth*APRT without any additional sequences. Nucleotide sequences of the cloned genes were verified by sequencing. The codone GGG→AGG substitution corresponding to amino-acid Arg27Gly replacement was found in the gene encoding the *Tth*HPRT.

The screening of available producer strains was performed to find strains, which produce target enzymes in soluble form. The resulting strains *E. coli* BL21(DE3)/pER- *Tth*APRT and *E. coli* C3030/pER- *Tth*HPRT produced enzymes mainly in soluble form (>80%).

The established procedure for isolation and purification of *Tth*HPRT includes heat treatment, immobilized metal affinity chromatography, final size-exclusion chromatography, and concentration. For *Tth*APRT, the protocol include heat treatment, anion exchange chromatography, hydrophobic chromatography, final size-exclusion chromatography, and concentration. Yields of both transferases were no less than 10–13 mg per liter of culture, with a purity of about 96% (as determined by SDS-PAGE).

The influence of temperature and Mg^{2+} concentration on the activity of *Tth*HPRT was investigated. The results were compared with data for adenine phosphoribosyltransferase *Thermus thermophilus*, obtained earlier [1].

The *Tth*APRT is active over a wide temperature range (Figure 3). A maximal activity of *Tth*HPRT (1.1 unit/mg) is observed at 60 °C. The activity at 36 °C is 5% from the maximal one and at 90 °C it is 3% from the maximal one. It is interesting, that *Tth*APRT shows its maximal activity at 75 °C.

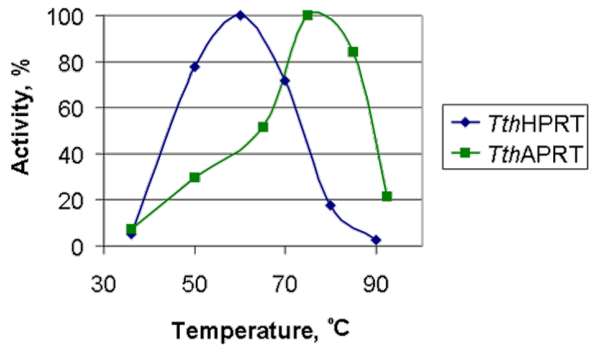


Figure 3: Dependence of *Tth*HPRT and *Tth*APRT activity on temperature (reaction mixtures (0.5 mL) contained 20 mM Tris-HCl, pH 8.0, 1 mM 5-phosphoribosyl-1-α-pyrophosphate, and 5 mM $MgCl_2$; in the case of *Tth*HPRT mixtures contained 1 mM hypoxanthine and 0.18 μg of enzyme, in the case of *Tth*APRT – 1 mM adenine and 0.125 μg of enzyme).

The influence of the magnesium ion concentration on the *Tth*HPRT activity is nonlinear. The activity increases rapidly while the magnesium chloride concentration increases from 0 to 1 mM (Figure 4). Further increasing of the concentration (up to 5 mM) does not increase the activity significantly. Since the reaction rate increases rapidly with increasing the magnesium chloride concentration to values equivalent to the concentration of 5-phosphoribosyl-α-1-pyrophosphate (1 mM), it can be assumed that the presence of magnesium ions promotes the proper spatial orientation of the substrate. The reaction also proceeds in the absence of magnesium ions in solution. A similar dependence is observed for *Tth*APRT.

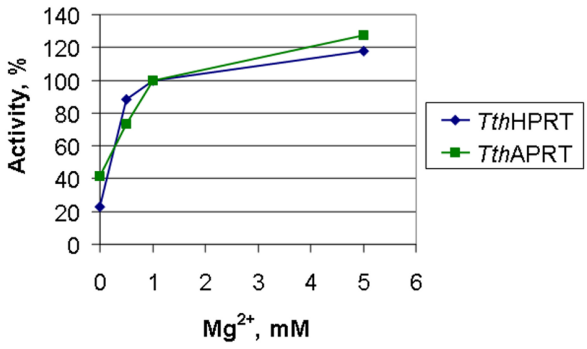


Figure 4: Dependence of *Tth*HPRT and *Tth*APRT activity on the Mg^{2+} concentration (reaction mixtures (0.5 mL) contained 20 mM Tris-HCl, pH 8.0, 1 mM 5-phosphoribosyl-1-α-pyrophosphate, and 0–5 mM $MgCl_2$; in case of *Tth*HPRT mixtures contained 1 mM hypoxanthine and 0.18 μg of enzyme, in case of *Tth*APRT – 1 mM adenine and 0.125 μg of enzyme; reactions were performed at 70 °C; 100% – activity at 1 mM concentration).

After optimization of the reaction conditions, kinetic parameters for *Tth*HPRT were determined (Table 2).

Based on the K_m values, the affinity of 5-phosphoribosyl-α-1-pyrophosphate for the active site is much lower than that of heterocyclic bases. The similar situation we observed for *Tth*APRT [1]. Comparison of the synthesis rates of inosine-5'-monophosphate and guanosine-5`-monophosphate showed that the first is synthesized 4.6 times faster. The literature data for similar enzymes (see Table 2) confirm a poor affinity of PRPP to the active site: K_m for hypoxanthine is 17 fold less then for PRPP, although for the human enzyme K_m is only 5 fold less. Comparing two enzymes from different strains of *Thermus thermophilus*, we can conclude that *Tth*APRT from HB8 (in contrast with HB27), synthesizes guanosine-5`-monophosphate faster. This may be due to the difference in reaction conditions. Kinetic data are displayed by double reciprocal plot (Figure 5). Determination of substrate specificity of *Tth*HPRT was performed in comparative experiments with *Tth*APRT. The process of nucleotide synthesis was monitored by a liquid chromatography–mass spectrometry analysis of the reaction mixture.

The data is presented in the Table 3. As expected, *Tth*HPRT is specific to 6-oxopurines, while *Tth*APRT is specific to 6-aminopurines. Both enzymes do not recognize thymine as a substrate. This is consistent with data that pyrimidine heterocyclic bases are substrates for uracyl phosphoribosyltransferase

Table 2: Kinetic parameters of inosine-5'-monophosphate and guanosine-5'-monophosphate synthesis using HPRT from various organisms.

Substrate	K_m, μM	V_{max}, μmol/min·mg	k_{cat}, 1/s	k_{cat}/K_m, 1/M·s
Thermus thermophilus HB27				
hypoxanthine	13 ± 4	28 ± 9	9 ± 3	6.9×10^5
guanine	28 ± 9	6 ± 2	2.0 ± 0.7	7.1×10^4
PRPP	220 ± 60	17 ± 5	6 ± 2	2.7×10^4
Thermus thermophilus HB8 [12]				
hypoxanthine	3.9 ± 1.5	–	9.1 ± 0.8	–
guanine	7.4 ± 1.7	–	18 ± 1	–
PRPP	68 ± 18	–	20 ± 2	–
Homo sapiens [13]				
hypoxanthine	3.8 ± 0.3	–	2.6 ± 0.6	7×10^5
PRPP	19.1 ± 1.6	–	2.5 ± 0.05	1×10^5
Escherichia coli [14]				
Hypoxanthine	37	–	–	–
PRPP	330	–	–	–

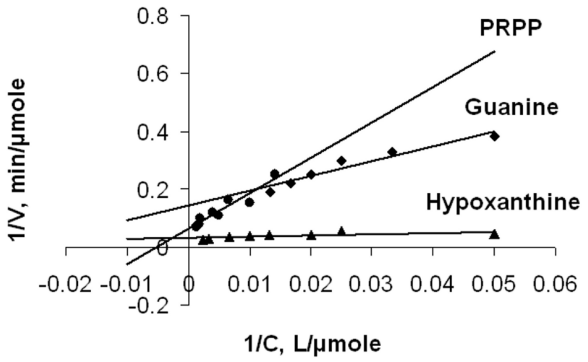


Figure 5: Lineweaver–Burk plot for synthesis of inosine-5'-monophosphate and guanosine-5'-monophosphate.

Table 3: Substrate specifity of *Tth*HPRT and *Tth*APRT.

Base	Conversion (24 h, %)[a]	
	*Tth*HPRT	*Tth*APRT
adenine	5.3	48.1
hypoxanthine	91.0	6.4
guanine	73.9	25.6
2-chloroadenine	0	52.9
2-fluoroadenine	0	31.1
6-mercapopurine	85.1	4.8
allopurinol	39.3	57.4
8-azaguanine	80.6	1.0
thymine	0	0
1,2,4-triazole-3-carboxamide	0	0
1,2,4-triazole-3-carboxy-*N*-methyl-amide	0	0

[a]Reaction mixtures (0.125 mL, 20 mM Tris-HCl, pH 8.0, 60 °C) contained 0.5 mM heterocyclic base, 0.5 mM PRPP, 0.5 mM $MgCl_2$, 5 mM KH_2PO_4 and 0.4 μg of *Tth*HPRT or *Tth*APRT.

and orotate phosphoribosyltransferase only [2]. Unfortunately, we did not find any product in reactions with compounds based on 1,2,4-triazole-3-carboxamide, which was also observed for *E. coli* HPRT [15,16]. However, allopurinol and 8-azaguanine are substrates for *Tth*HPRT, and 2-chloroadenine is a substrate for *Tth*APRT. For 2-chloroadenine and 8-azaguanine, reaction at a higher temperature is preferable because of their low solubility in water (less than 1 mM at 37 °C). Interestingly, allopurinol proved to be a good substrate for both *Tth*HPRT and *Tth*APRT, unlike hypoxanthine, which differs only in the position of one of the nitrogen atoms. Probably, the presence of nitrogen atom at C7 position of purine heterocycles plays an important role in reactions catalyzed by phosphoribosyltransferase, and also affects the substrate properties of *Tth*HPRT and *Tth*APRT.

Two nucleotides were synthesized using *Tth*HPRT or *Tth*APRT (see Figure 6). Synthesis of 2-Cl-AMP was performed at 75 °C. This allowed to achieve a concentration of 0.5 mM of the initial 2-chloroadenine. The reaction progress was monitored by HPLC. After 2 days (the product content in the reaction mixture was 54%), the reaction mixture was concentrated and the desired product was isolated by column chromatography on ion-exchange sorbents (anion and then cation-exchange). The yield of 2-Cl-AMP was 37%. A second nucleotide (Allop-MP) was synthesized at a lower temperature (60 °C). After 2 days, the product content in the reaction mixture was 55%. The product was isolated in the same way, with a yield of 32%.

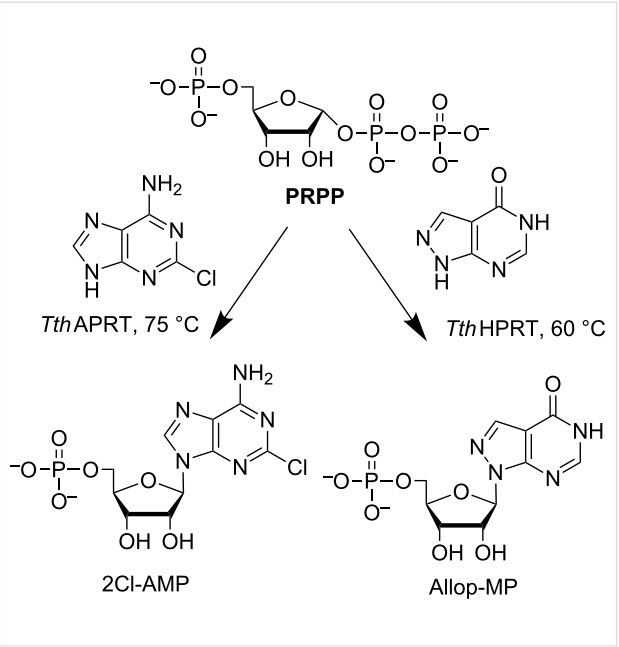


Figure 6: Synthesis of nucleotides 2Cl-AMP and Allop-MP using phosphoribosyltransferases *Tth*APRT or *Tth*HPRT respectively.

Conclusion

The recombinant adenine phosphoribosyltransferase and hypoxanthine phosphoribosyltransferase from *Thermus thermophilus* HB27 were purificated with yields no less than 10–13 mg per litre of culture. A comparative study of substrate specificity of these enzymes towards different heterocyclic bases was carried out and temperature-dependence and magnesium chloride concentration-dependence of enzymes activity were determined. *Tth*HPRT can be used for the synthesis of nucleotides containing different purine derivatives including 8-aza- and 8-aza-7-deazapurine. The use of hypoxanthine and adenine transferases in multi-enzyme cascades significantly extends the spectrum of synthetic purine nucleotides. Two nucleotides were synthesized: 9-(β-D-ribofuranosyl)-2-chloroadenine 5'-monophosphate (2-Cl-AMP) using *Tth*APRT and 1-(β-D-ribofuranosyl)pyrazolo[3,4-*d*]pyrimidine-4-one 5'-monophosphate (Allop-MP) using *Tth*HPRT with yields of 37% and 32%, respectively. Using of hypoxanthine and adenine transferases in multi-enzyme cascades significantly extends the spectrum of synthetic purine nucleotides.

Experimental

Tris-buffer, acetic acid, sodium chloride, glycerol, acrylamide, *N*,*N*'-bisacrylamide, ATP, bromophenol blue, agarose, EDTA, IPTG, ampicillin, sodium dodecylsulfate, imidazole and DMF were purchased from Panreac (Spain, Barselona). Ethanol was purchased from MedChemProm. Coomassie Brilliant Blue R-250 was purchased from Bio-Rad (USA, CA). Bacto yeast extract, bacto tryptone, and bacto agar were purchased from Becton Dickinson Biosciences (USA, NJ). NaOH and HCl were purchased from Merck (USA, MA). Sodium persulfate, TEMED, ethidium bromide, and sodium azide were purchased from Helicon (Russia). dNTP was purchased from Thermo Fisher Scientific (USA, MA). DTT, phenylmethylsulfonyl chloride, magnesium chloride, nickel sulfate, potassium dihydroorthophosphate, Ni-IDA sepharose, 5-phosphoribosyl-α-1-pyrophosphate and all bases (adenine, hypoxanthine, guanine, 2-chloroadenine, etc.) were purchased from Sigma-Aldrich (USA, MO).

Bacterial strains: a) *E. coli* C3030 [MiniF *lysY* (Cam^R) / *fhuA2 lacZ::T7 gene1* [lon] *ompT ahpC gal λatt*::pNEB3-r1-*cDsbC* ($Spec^R$, $lacI^q$) *ΔtrxB sulA11 R(mcr-73::miniTn10*--Tet^S)2 [dcm] *R(zgb-210::Tn10* --Tet^S) *endA1 Δgor* Δ*(mcrC-mrr)114::IS10*] New England Biolabs (USA, MA), b) *E.coli* BL21(DE3) *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*.

Plasmid vector: pET 23a+, pET 23d+ (Merck Millipore, USA, MA).

Enzymes: NdeI, XhoI, NcoI, T4 DNA-ligase (Thermo Scientific, USA, MA), Encyclo-polymerase (Eurogen, Russia).

The protein concentration was determined by the Bradford method [17], using BSA as a standard.

Protein purity was determined by electrophoresis in a polyacrylamide gel under denaturing conditions [18].

Cloning and creation of producer strain: Genes TT_RS08985 and TT_RS06315, encoding *Tth*HPRT and *Tth*APRT, respectively, were amplified on the genomic DNA template of the *T. thermophilus* HB27 strain by a polymerase chain reaction (PCR) using synthetic primers. The genes were cloned into the expression vectors pET-23a+ and pET-23d+ respectively. The *E. coli* strains BL21(DE3)/pER- *Tth*APRT and C3030/pER- *Tth*HPRT produced the target enzymes mainly in soluble form (culturing conditions: 4 h grow at 37 °C after supplementing with 0.4 mM IPTG).

Isolation and purification of *Tth*HPRT: A cell pellet was resuspended in 50 mM Tris-HCl, 200 mM NaCl, and 1 mM phenylmethylsulfonyl fluoride (PMSF) pH 8.0 (1:10 w/v). The cells were disrupted by sonication for 30 min at 20 kHz at +4 °C. The cell debris was pelleted by centrifugation at 12,000 rpm for 30 min at +4 °C. The cell lysate was heat-treated (65 °C, 10 min) and the pellet was removed by centrifugation. The supernatant was applied to a column XK 16/20 (GE Healthcare, USA) packed with Ni-IDA Sepharose (Sigma Aldrich,

USA) pre-equilibrated with 50 mM Tris-HCl and 200 mM NaCl at pH 8.0. Ballast proteins were eluted with a solution, containing 50 mM Tris-HCl, 50 mM imidazole, and 200 mM NaCl, pH 8.0 (4 CV, flow rate 2 mL/min). The target enzyme was eluted with solution, contained 50 mM Tris-HCl, 250 mM imidazole, and 200 mM NaCl, pH 8.0 (4 CV, flow rate 2 mL/min). Pooled fractions were concentrated by a polysulfonic membrane PBGC 10 kDa (Millipore, USA). The resulting solution was applied to a column with HiLoad 16/60 Superdex 75pg (GE Healthcare, USA), equilibrated by 20 mM Tris-HCl, 50 mM NaCl, 0.04% NaN_3, and 10% glycerol, pH 8.0. Fractions, containing the target enzyme with a purity of more than 96 %, were pooled and concentrated up to a concentration of 13 ± 1 mg/mL.

Isolation and purification of *Tth*APRT: Cell biomass disruption and heat-treatment was performed as described in section "Isolation and purification of *Tth*HPRT". The resulting solution was diluted (2-fold) with solution, comtained 50 mM Tris-HCl, 2 M $(NH_4)_2SO_4$, pH 8.0, and applied to column XK 16/20 packed with Phenyl Sepharose HP (GE Healthcare, USA). The column was eluted by linear gradient of $(NH_4)_2SO_4$ (1.0 – 0 M, 12 CV, flow rate 2 mL/min). Fractions, contained the target enzyme, were pooled and concentrated on polysulphonic membrane PBGC 10 kDa. The resulting solution was applied to column with HiLoad 16/60 Superdex 200, equilibrated by 20 mM Tris-HCl, 50 mM NaCl, 0.04% NaN_3, and 5% glycerol, pH 8.0. Fractions, contained the target enzyme with purity more than 96%, were pooled and concentrated up to concentration 12 ± 1 mg/mL.

Enzyme assay: Each reaction mixture (0.5 mL, 20 mM Tris-HCl, pH 8.0) contained 1 mM 5-phosphoribosyl-1-α-pyrophosphate, 1 mM hypoxanthine, 5 mM $MgCl_2$, and hypoxanthine phosphoribosyltransferase *Thermus thermophilus* (0.18 μg). Reaction mixtures were incubated at 70 °C. Substrate and product quantities were determined using HPLC (Waters 1525, column Ascentis Express C18, 2.7 μm, 3.0 × 75 mm, eluent A 0.1% aqeous TFA, eluent B 0.1% TFA / 70% acetonitrile in water, detection at 254 nm, Waters 2489).

Kinetic parameters determination: Each reaction mixture (1.0 mL, 20 mM Tris-HCl, pH 8.0) contained 5 mM $MgCl_2$, hypoxanthine phosphoribosyltransferase *Thermus thermophilus* (0.18 μg), and the following components: a) hypoxanthine (0.01–0.50 mM) or guanine (0.01–0.20 mM) and 1 mM 5-phosphoribosyl-1-α-pyrophosphate to determine K_m and V_{max} for hypoxanthine and guanine, and b) 5-phosphoribosyl-1-α-pyrophosphate (0.05–1.20 mM) and 0.50 mM hypoxanthine to determine K_m and V_{max} for 5-phosphoribosyl-1-α-pyrophosphate. Reaction mixtures were incubated at 70 °C for 2 min. Product quantities were determined as described in the "Enzyme assay" section. Each experiment was repeated three times. Kinetic parameters were determined by nonlinear regression analysis using SciDAVis v0.2.4 software (free software, web site: scidavis.sourceforge.net). Catalytic constants (k_{cat}) were calculated per 1 subunit (20.3 kDa, calculated based on amino acid sequence).

Mass spectra were measured on an Agilent 6224, ESI-TOF, LC/MS (USA) in positive ion mode (ESI), LCQ Fleet ion trap mass spectrometer (Thermo Electron, USA) and Agilent 1100 LC/MSD VL (Agilent Technologies) equipped an APCI and ESI source (positive and negative mode of ionization), 1100 DAD and ELSD PL-ELS 1000 (Polymer Laboratories).

Nucleotides synthesis

9-(β-D-Ribofuranosyl)-2-chloroadenine 5'-monophosphate (2-Cl-AMP): 2-Chloroadenine (17 mg, 0.10 mmol) was dissolved in water (203 mL) under stirring and heating at 90 °C , and after cooling to 70 °C, magnesium chloride hexahydrate (41 mg, 0.21 mmol) and potassium dihydroorthophosphate (276 mg, 2.03 mmol) were added. The pH of the solution was adjusted to 8.0 by 2 N potassium hydroxide. The pentasodium salt of 5-phosphoribosyl-α-1-pyrophosphate (70 mg, 0.14 mmol) and *Tth*APRT (5 units) were added, and the reaction mixture was incubated at 75 °C for 2 days; the reaction progress was monitored by HPLC. The reaction mixture was neutralized with 2 N hydrochloric acid and concentrated in vacuo to ca. 10 mL. The precipitate was filtered off, the filtrate was applied to the column with DEAE-Toyopearl 650C, bicarbonate form, 40 × 140 mm, and the product was eluted with triethylammonium bicarbonate (0.1 M). Fractions were concentrated in vacuo to ca. 10 mL, applied to the column with CM-Sephadex C-25, sodium form, 20 × 160 mm, and the product was eluted with water to give, after evaporation and drying in vacuo under P_2O_5, 16 mg (0.037 mmol; 37%) of 9-(β-D-ribofuranosyl)-2-chloroadenine 5'-monophosphate of 99% purity (HPLC). HRMS (ESI^+): *m/z* $[M + H]^+$ calcd for $C_{10}H_{13}N_5O_7P_1Cl_1$: 382.0315; found, 382.0353; $[2M + H]^+$, found, 763.0606; $[Base + H]^+$, found, 170.0244; ^1H NMR (DMSO-d_6) δ ppm) 8.52 (s, 1H, H8), 7.78 (br. s., 2H, NH_2), 5.83 (d, $J_{1',2'}$ = 6 Hz, 1H, H1'), 4.61 (m, 1H, H2'), 4.23 (m, 1H, H3'), 4.06 (m, 1H, H4'), 3.84 (m, 2H, H5a', H5b') ppm; ^{13}C NMR (DMSO-d_6) δ 156.61 (C2 or C6), 153.01 (C6 or C2), 150.73 (C4), 139.68 (C8), 117.57 (C5), 86.40 (C1'), 84.52 (C4'), 74.13 (C2'), 71.06 (C3'), 3.94 (C5') ppm; ^{15}N NMR (DMSO-d_6) δ 242.7 (N7), 171.3 (N9), 86.84 (NH_2) ppm.

1-(β-D-Ribofuranosyl)pyrazolo[3,4-*d*]pyrimidine-4-one 5'-monophosphate (Allop-MP): Allopurinol (14 mg,

0.10 mmol) was dissolved in water (203 mL) under stirring and heating at 90 °C, and after cooling to 50 °C, magnesium chloride hexahydrate (41 mg, 0.21 mmol) and potassium dihydroorthophosphate (276 mg, 2.03 mmol) were added. The pH of the solution was adjusted to 8.0 with 2 N potassium hydroxide. The pentasodium salt of 5-phosphoribosyl-α-1-pyrophosphate (70 mg, 0.14 mmol) and *Tth*HPRT (5 units) were added, and the reaction mixture was incubated at 60 °C for 2 days; the reaction progress was monitored by HPLC. The reaction mixture was neutralized by 2 N hydrochloric acid and concentrated in vacuo to ca. 10 mL. The precipitate was filtered off, the filtrate was placed on the column with DEAE-Toyopearl 650C, bicarbonate form, 40 × 140 mm, and the product was eluted with triethylammonium bicarbonate (0.2 M). Fractions were concentrated in vacuo to ca. 10 mL, applied to the column with CM-Sephadex C-25, sodium form, 20 × 160 mm, and the product was eluted with water to give, after evaporation and drying in vacuo under P_2O_5, 11 mg (0.032 mmol; 32%) of 1-(β-D-ribofuranosyl)pyrazolo[3,4-*d*]pyrimidine-4-one 5'-monophosphate of 97% purity (HPLC). HRMS (ESI$^+$): *m/z* $[M + H]^+$ calcd for $C_{10}H_{13}N_4O_8P_1$: 349.0545; found, 349.0520; $[2M + H]^+$, found, 697.0952; $[3M + H]^+$, found, 1045.1374; $[Base + H]^+$ found, 137.0453; ^{1}H NMR (DMSO-d_6) δ 12.44 (br. s, 1H, NH), 8.15 (s, 1H, H3), 8.13 (s, 1H, H6), 6.06 (d, *J* = 4.1 Hz, 1H, H1'), 4.56 (dt, 1H, H2', *J* = 4.54; <0.5), 4.31 (t, 1H, H3', *J* = 4.8), 4.04 (m, 1H, H4'), 3.85 (ddd, *J* = 11.0, 7.6; 6.2 Hz, 1H, H5'a), 3.66 (ddd, *J* = 11.0, 7.2, 6.1 Hz 1H, H5'b) ppm; ^{13}C NMR (DMSO-d_6) δ 157.03 (C4), 152.90 (C7a), 148.53 (C6), 135.38 (C3), 106.06 (C4a), 88.16 (C1'), 83.27 (C4'), 73.39 (C2'), 71.38 (C3'), 64.76 (C5') ppm; ^{15}N NMR (DMSO-d_6) δ 302.8 (N2), 210.6 (N7), 204.9 (N1), 171.1 (N5).

Abbreviations

APRT – adenine phosphoribosyltransferase; HPRT – hypoxathine phosphoribosyltransferase; PRPPS – phosphoribosylpyrophosphate synthetase; RK – ribokinase; *Tth* - *Thermus thermophilus;* 2-Cl-AMP – 9-(β-D-ribofuranosyl)-2-chloroadenine 5'-monophosphate; Allop-MP – 1-(β-D-ribofuranosyl)-pyrazolo[3,4-*d*]pyrimidine-4-one 5'-monophosphate

Supporting Information

Supporting Information File 1

Detailed analysis of mass spectrometry and NMR data.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-14-289-S1.pdf]

Acknowledgements

The authors are grateful to Russian Science Foundation (project No. 14-50-00131) for financial support of this work.

ORCID® iDs

Ilja V. Fateev - https://orcid.org/0000-0002-8571-4197
Irina D. Konstantinova - https://orcid.org/0000-0001-5563-6549

References

1. Esipov, R. S.; Abramchik, Yu. A.; Fateev, I. V.; Konstantinova, I. D.; Kostromina, M. A.; Muravyova, T. I.; Artemova, K. G.; Miroshnikov, A. I. *ActaNaturae* **2016,** *8,* 82–90.
2. Del Arco, J.; Fernández-Lucas, J. *Curr. Pharm. Des.* **2017,** *23,* 1–15. doi:10.2174/1381612823666171017165707
3. Taran, S. A.; Verevkina, K. N.; Feofanov, S. A.; Miroshnikov, A. I. *Russ. J. Bioorg. Chem.* **2009,** *35,* 739–745. doi:10.1134/s1068162009060107
4. Mikhailopulo, I. A. *Curr. Org. Chem.* **2007,** *11,* 317–335. doi:10.2174/138527207780059330
5. Mikhailopulo, I. A.; Miroshnikov, A. I. *Mendeleev Commun.* **2011,** *21,* 57–68. doi:10.1016/j.mencom.2011.03.001
6. Wintersberger, E. *Biochem. Soc. Trans.* **1997,** *25,* 303–308. doi:10.1042/bst0250303
7. Tesmer, J. J.; Klem, T. J.; Deras, M. L.; Davisson, V. J.; Smith, J. L. *Nat. Struct. Biol.* **1996,** *3,* 74–86. doi:10.1038/nsb0196-74
8. Kim, M.-J.; Whitesides, G. M. *Appl. Biochem. Biotechnol.* **1987,** *16,* 95–108. doi:10.1007/bf02798359
9. Scism, R. A.; Stec, D. F.; Bachmann, B. O. *Org. Lett.* **2007,** *9,* 4179–4182. doi:10.1021/ol7016802
10. Mikhailopulo, I. A.; Miroshnikov, A. I. *ActaNaturae* **2010,** *2,* 36–59.
11. Iglesias, L. E.; Lewkowicz, E. S.; Medici, R.; Bianchi, P.; Iribarren, A. M. *Biotechnol. Adv.* **2015,** *33,* 412–434. doi:10.1016/j.biotechadv.2015.03.009
12. Kanagawa, M.; Baba, S.; Ebihara, A.; Shinkai, A.; Hirotsu, K.; Mega, R.; Kim, K.; Kuramitsu, S.; Sampei, G.-i.; Kawai, G. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2010,** *66,* 893–898. doi:10.1107/s1744309110023079
13. Canyuk, B.; E-Wan, A.; Keawwijit, W.; Nualnoi, T.; Sirisatean, L.; Tansakul, P.; Tanthana, C. *Nucleosides, Nucleotides Nucleic Acids* **2008,** *27,* 894–899. doi:10.1080/15257770802146593
14. Miller, R. L.; Ramsey, G. A.; Krenitsky, T. A.; Elion, G. B. *Biochemistry* **1972,** *11,* 4723–4731. doi:10.1021/bi00775a014
15. Streeter, D. G.; Miller, J. P.; Robins, R. K.; Simon, L. N. *Ann. N. Y. Acad. Sci.* **1977,** *284,* 201–210. doi:10.1111/j.1749-6632.1977.tb21952.x
16. Naesens, L.; Guddat, L. W.; Keough, D. T.; van Kuilenburg, A. B. P.; Meijer, J.; Voorde, J. V.; Balzarini, J. *Mol. Pharmacol.* **2013,** *84,* 615–629. doi:10.1124/mol.113.087247
17. Laemmli, U. K. *Nature* **1970,** *227,* 680–685. doi:10.1038/227680a0
18. Bradford, M. M. *Anal. Biochem.* **1976,** *72,* 248–254. doi:10.1016/0003-2697(76)90527-3

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

New standards for collecting and fitting steady state kinetic data

Kenneth A. Johnson

Review

Open Access

Address:
Department of Molecular Biosciences, The University of Texas, Austin, TX 78735, USA

Email:
Kenneth A. Johnson - kajohnson@mail.uexas.edu

Keywords:
computer simulation; data fitting; enzyme catalysis; induced-fit; Michaelis constant; specificity constant

Beilstein J. Org. Chem. **2019,** *15,* 16–29.
doi:10.3762/bjoc.15.2

Received: 22 October 2018
Accepted: 04 December 2018
Published: 02 January 2019

This article is part of the thematic issue "Enzymes in chemical transformations".

Guest Editor: K. N. Allen

Abstract

The Michaelis–Menten equation is usually expressed in terms of k_{cat} and K_m values: $v = k_{cat}[S]/(K_m + [S])$. However, it is impossible to interpret K_m in the absence of additional information, while the ratio of k_{cat}/K_m provides a measure of enzyme specificity and is proportional to enzyme efficiency and proficiency. Moreover, k_{cat}/K_m provides a lower limit on the second order rate constant for substrate binding. For these reasons it is better to redefine the Michaelis–Menten equation in terms of k_{cat} and k_{cat}/K_m values: $v = k_{SP}[S]/(1 + k_{SP}[S]/k_{cat})$, where the specificity constant, $k_{SP} = k_{cat}/K_m$. In this short review, the rationale for this assertion is explained and it is shown that more accurate measurements of k_{cat}/K_m can be derived directly using the modified form of the Michaelis–Menten equation rather than calculated from the ratio of k_{cat} and K_m values measured separately. Even greater accuracy is achieved with fitting the raw data directly by numerical integration of the rate equations rather than using analytically derived equations. The importance of fitting to derive k_{cat} and k_{cat}/K_m is illustrated by considering the role of conformational changes in enzyme specificity where k_{cat} and k_{cat}/K_m can reflect different steps in the pathway. This highlights the pitfalls in attempting to interpret K_m, which is best understood as the ratio of k_{cat} divided by k_{cat}/K_m.

Review

When Henri, Michaelis and Menten derived the equation for steady state enzyme turnover, they chose to define the rate in terms of V_{max} and the substrate dissociation constant for the hypothetical enzyme–substrate complex, K_S [1,2].

$$V = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_S + [S]}$$

$$V_{max} = k_{cat} \cdot [E]_0$$

At the time, the choice of the terms V_{max} and K_S was logical because the concentrations of enzymes could not be determined and even the chemical makeup of enzymes was unknown. By including the unknown enzyme concentration in the term for the maximum velocity of turnover, the equation contained two variables, V_{max} and K_S, consistent with the information content of the data and a minimal model.

In 1913, Michaelis and Menten provided evidence for the existence of an enzyme–substrate complex by careful rate measurements and rigorous quantitative analysis, fulfilling the major goal of their work [1,2]. Estimating the binding affinity for the substrate as K_S was an added bonus. These were profound discoveries that laid the foundation for enzymology throughout the 20th century.

The Michaelis–Menten equation was originally derived assuming that substrate binding was at equilibrium, and was later expanded by Briggs and Haldane [3] who used the steady state approximation to include the rates of substrate and product release in defining K_m according to a minimal model.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

$$K_m = \frac{k_{-1} + k_2}{k_1}$$

A century later, we know the structures of enzymes and can accurately determine their concentrations so we divide the measured rates by the known enzyme concentration to get the common form of the Michaelis–Menten equation:

$$v = \frac{d[P]/dt}{[E]_0} = \frac{k_{cat}[S]}{K_m + [S]} \quad (1)$$

Using this equation, the two parameters derived in fitting data are k_{cat} and K_m, from which we can calculate k_{cat}/K_m. However, k_{cat}/K_m is the most important parameter as it is used to quantify enzyme specificity, efficiency and proficiency [4,5]. In fact, k_{cat} and k_{cat}/K_m should be considered as the two primary steady state kinetic parameters, rather than k_{cat} and K_m. A half century ago Cleland stressed that the two fundamental steady state kinetic parameters were k_{cat} and k_{cat}/K_m and that K_m represents an "apparent dynamic dissociation constant under steady state conditions", but noted that K_m is not an independent parameter [6]. This statement was based on the use of a Lineweaver–Burk (double-reciprocal) plot [7] to fit data where the intercept defines $1/k_{cat}$ and the slope defines $1/k_{cat}/K_m$.

$$\frac{1}{v} = \frac{1}{k_{cat}} + \frac{K_m}{k_{cat}} \cdot \frac{1}{[S]}$$

In Cleland's analysis, the two primary steady state kinetic parameters were k_{cat} and k_{cat}/K_m because they were the parameters derived in fitting data displayed on a double reciprocal plot. Today, the emphasis is on interpreting the steady state kinetic parameters in terms of enzyme structure and individual steps in the reaction pathway. This leads to a new justification for choosing k_{cat}/K_m rather than K_m as a primary kinetic parameter. Of the three steady state parameters (k_{cat}, K_m, and k_{cat}/K_m) k_{cat}/K_m is the most important as it quantifies enzyme specificity, efficiency and proficiency [4]. Moreover, both k_{cat} and k_{cat}/K_m place lower limits on rate constants for individual steps in the pathway, while K_m is largely un-interpretable.

Cleland published the first computer programs [8] to fit data based on linear regression of data displayed on a double-reciprocal plot, and including a kind of global analysis in resolving steady state inhibition patterns, which are defined by the effects of inhibitors on the slope and intercept, i.e., k_{cat} and k_{cat}/K_m. However, there are serious disadvantages in using a double reciprocal plot due to the unequal weighting of errors and the compression of the most accurate data leading to a distorted view of the results. The unequal weighting of errors can be overcome if the standard deviations of the individual measurements are included in the linear regression analysis, but that is not always done.

Regardless of the method used to fit data, there is merit in fitting to derive k_{cat} and k_{cat}/K_m, rather than fitting to derive k_{cat} and K_m individually then calculating k_{cat}/K_m from the ratio. There are large errors in k_{cat} and K_m since these estimates each rely on extrapolation to infinite concentration of substrate, leading to larger errors in the calculated k_{cat}/K_m value. On the other hand, the value of k_{cat}/K_m is generally well defined from the initial slope of the concentration dependence, as illustrated in Figure 1. Thus, k_{cat}/K_m can be understood as the apparent second order rate constant for substrate binding. More precisely, k_{cat}/K_m is equal to the true second order rate constant for substrate binding to the enzyme multiplied by the probability that the bound substrate will be converted to product and released into solution. This principle can be illustrated using the simplest model:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

$$k_{cat}/K_m = k_1 \cdot \frac{k_2}{k_{-1} + k_2}$$

The term, $k_2/(k_{-1} + k_2)$, gives the probability that the substrate reacts rather than dissociating. With more realistic models, the more complex equations for k_{cat}/K_m can still be understood as the product of the rate constant for substrate binding times the probability of forward reaction.

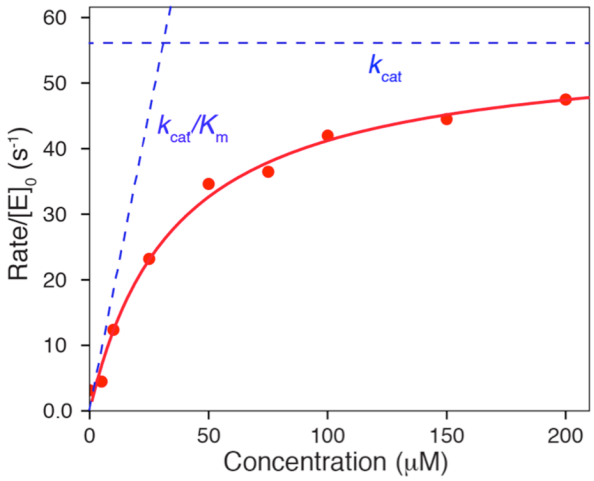


Figure 1: Michaelis–Menten plot. The rate of product formation is plotted versus substrate concentration and fit to a hyperbola. The dashed lines illustrate k_{cat}/K_m (slope) and k_{cat}. The intersection of the two lines gives the substrate concentration at which k_{cat}/K_m $[S]_i = k_{cat}$, so $[S]_i = k_{cat}/(k_{cat}/K_m) = K_m$.

Interpretation of steady state kinetic parameters takes on new significance in the current era of enzymology where the emphasis is on relating the parameters to individual rate constants and to structural and chemical transitions for each reaction in the pathway. While k_{cat}/K_m can be directly interpreted in terms of enzyme specificity, it also provides a lower limit for the second order rate constant for substrate binding. Similarly, k_{cat} provides a lower limit for each first order rate constant following substrate binding through product release. On the other hand, the Michaelis constant cannot be interpreted unambiguously in the absence of additional information. In fact, K_m can be less than, greater than, or equal to the K_d for substrate binding. Here, the overly simplified model gives the wrong answer in suggesting that K_m is always greater than or equal to the dissociation constant (K_d).

$$\text{If } K_m = \frac{k_{-1} + k_2}{k_1} \text{ then } K_m \geq K_d = \frac{k_{-1}}{k_1}$$

However, for a more complete model we come to a different conclusion:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \xrightarrow{k_3} E + P$$

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_{-2} + k_3}$$

$$K_m = \frac{k_2 k_3 + k_{-1}(k_{-2} + k_3)}{k_1(k_2 + k_{-2} + k_3)} \quad (2)$$

$$k_{cat}/K_m = k_1 \frac{k_2 k_3}{k_2 k_3 + k_{-1}(k_{-2} + k_3)}$$

We can now see that depending on the intrinsic rate constants, K_m can be less than, greater than, or equal to the K_d. Thus, in the absence of additional information, K_m cannot be interpreted to imply anything about the intrinsic rate and equilibrium constants governing enzyme catalysis. Although the K_m defines the concentration of substrate giving half maximal velocity, that is a phenomenological description without any mechanistic implications. On the other hand, k_{cat} and k_{cat}/K_m provide meaningful lower limits on intrinsic rate constants.

The best understanding of K_m is as the ratio of k_{cat} and k_{cat}/K_m, so we consider that the Michaelis constant is a derivative of the two primary steady state kinetic parameters.

$$K_m = \frac{k_{cat}}{k_{cat}/K_m}$$

Although this statement appears as trivial algebra, it is profound because k_{cat} and k_{cat}/K_m can reflect different steps in the enzyme pathway as will be described below. Moreover, it shows that the K_m value represents the balance point between the rate of turnover and the rate of substrate binding. The K_m represents substrate binding affinity only in the special case of rapid equilibrium binding.

A primary goal of fitting steady state data should be to accurately define k_{cat}/K_m. Rather than fitting to obtain estimates for k_{cat} and K_m and then calculating k_{cat}/K_m as a ratio, a more accurate view is to consider k_{cat} and k_{cat}/K_m as the primary steady state constants while K_m is obtained from their ratio. Traditionally, data have been fit using the standard form of the Michaelis–Menten equation to derive estimates of k_{cat} and K_m which are then used to calculate k_{cat}/K_m. However, there are often large errors in k_{cat} and K_m because these parameters are based on extrapolation to infinite substrate concentration, and these errors are compounded in calculating k_{cat}/K_m. Thus it is better to fit the data using an equation that provides k_{cat}/K_m directly using the following form of the Michaelis–Menten equation:

$$v = \frac{k_{SP}[S]}{1 + k_{SP}[S]/k_{cat}} \quad \text{where: } k_{SP} = k_{cat}/K_m \quad \text{or rather: } K_m = k_{cat}/k_{SP} \tag{3}$$

We use the term $k_{SP} = k_{cat}/K_m$ to emphasize that the specificity constant (k_{SP}) is a single parameter rather than a ratio and to stress that it represents the apparent second order rate constant for substrate binding. The use of the new term, k_{SP}, overcomes the awkward use of k_{cat}/K_m, which is not only more difficult to say and write, but it also presents the mistaken impression that it is simply a function of the rate of enzyme turnover divided by the substrate binding affinity. The awkwardness is the result of historical precedent. Defining the specificity constant as k_{cat}/K_m carries with it the baggage of thinking of the specificity constant as a ratio rather than a single parameter. Logic is influenced by the words we use to describe observations. It actually might help to avoid confusion in interpretation of K_m if we referred to the Michaelis constant as k_{cat}/k_{SP}.

Measuring k_{cat}/K_m

In order to get the best estimates of k_{cat}/K_m from steady state kinetic data, it is preferable to fit the data to Equation 3 in which k_{cat} and k_{cat}/K_m are the two fitted parameters rather than k_{cat} and K_m (Equation 1). To test this assertion, synthetic data were generated by computer simulation with k_{cat} = 50 s^{-1} and K_m = 20 μM. Data were generated at various concentrations of substrate (5, 10, 20, 30, and 40 μM), with a Gaussian distribution of random noise added to the data with a standard deviation of 0.5. The synthetic data were then fit to a straight line to estimate the rate (Figure 2A), which was then plotted versus substrate concentration and fit by nonlinear regression using either Equation 1 or Equation 3, defining k_{cat} and K_m or k_{cat} and k_{cat}/K_m, respectively.

As shown in Figure 2B,C, the fitted curves derived from either equation were indistinguishable, but as shown in Table 1 the error estimates in the fitted parameters were markedly different. The known standard deviation (sigma) values were included in the linear regression to estimate the rates and then the standard error estimates in fitted parameters were propagated to yield error estimates in k_{cat}/K_m. That is in computing z from the ratio of x and y, we compute the errors according to:

$$z = \frac{x}{y}$$

$$\frac{dz}{z} = \sqrt{\left(\frac{dx}{x}\right)^2 + \left(\frac{dy}{y}\right)^2},$$

where dx, and dy represent the error estimates on the variables x and y, respectively.

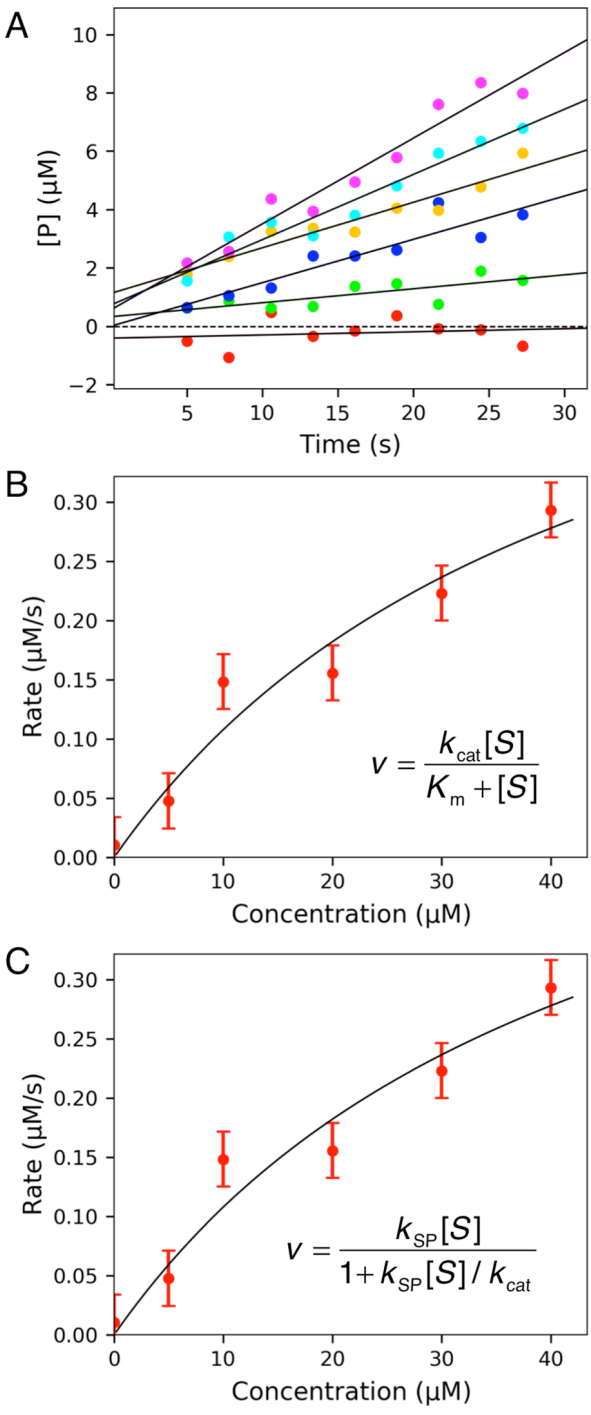


Figure 2: Fitting data to derive k_{cat} and K_m. A) Synthetic data were fit to a straight line and then the observed rate was plotted versus substrate concentration (B, C). In B), the data were fit to standard Michaelis–Menten equation while in C), the data were fit to a modified form equation. The equations used are shown in each figure.

Table 1: Summary of fitted parameters. Synthetic data were generated as described in the text and then fit to derive estimates of steady state kinetic parameters using different equations and different means of data fitting. Here we list the methods for fitting and the chosen fitted parameters for each method. The values for k_{cat}, K_m and k_{cat}/K_m are shown. Values in brackets were calculated from the other parameters. For example, in the first row, k_{cat}/K_m was calculated from k_{cat} and K_m, whereas in the second row, K_m was calculated from estimates of k_{cat} and k_{cat}/K_m. Standard error estimates for calculated parameters were obtained by propagation of errors as described in the text. Note that the fitted parameters need not reproduce the input parameters used to generate synthetic because of the added errors and the limited set of data. A more complete data set with lower errors would return the input parameters.

method	fitted parameters	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)
equations	k_{cat} and K_m	58.6 ± 25	44.4 ± 31	[1.32 ± 1.08]
	k_{cat} and k_{SP}	58.7 ± 25	[44.4 ± 23]	1.32 ± 0.37
simulation	k_{cat} and K_m	57.2 ± 5.9	27.2 ± 5.1	[2.11 ± 0.45]
	k_{cat} and k_{SP}	57.3 ± 6.1	[27.8 ± 4.1]	1.62 ± 0.21
simulation full reaction	k_{cat} and K_m	54.3 ± 4.9	23.6 ± 4.0	[2.30 ± 0.44]
	k_{cat} and k_{SP}	55.2 ± 3.9	[24.9 ± 2.2]	2.22 ± 0.12
input values		50	20	2.5

Table 1 illustrates the improvements in error estimates when fitting the data to derive k_{cat}/K_m (k_{SP}) directly rather than calculating the value from the ratio of k_{cat} and K_m. This is due to the large errors in estimating k_{cat} and K_m which are both based on extrapolation to infinite substrate concentration. In essence, the extrapolation errors are counted twice since they are reflected in both k_{cat} and K_m values. In contrast, when fitting to derive k_{cat} and k_{cat}/K_m, only k_{cat} is based on extrapolation while $k_{SP} = k_{cat}/K_m$ is obtained from the initial slope of the concentration dependence of the measured rate (Figure 1).

Admittedly, the "experiment" was set up to provide data only up to twice the K_m value to mimic those situations where the substrate concentrations available for testing are limited, so the exercise may not accurately reflect all laboratory settings. In that sense, the example may be biased in favor of fitting to derive k_{cat}/K_m directly. However, as a counterpoint, the only "experimental errors" in the data are random since the added noise conforms to a normal distribution, so this may make the fitting to define k_{cat} and K_m more accurate than seen in the laboratory. The "experiment" was repeated three times by generating new synthetic data and then fitting the data to derive independent k_{cat} and K_m values. The averages from this analysis were k_{cat} = 45.4 ± 15.9 s^{-1} and K_m = 33.2 ± 18.3 μM, which give k_{cat}/K_m = 1.36 ± 0.9 $\mu M^{-1}s^{-1}$. Averaging multiple experiments did not help to reduce errors as much as simply fitting data to a better equation.

One could argue that the choice of which equation to use is somewhat arbitrary. However, the common form of the equation was chosen over one hundred years ago for reasons that are no longer valid. Therefore, this historical precedent should no longer be followed given the advantage of fitting data to an equation that affords k_{cat}/K_m directly.

Fitting by simulation

Significant errors are introduced when fitting the primary data (product versus time) to a straight line because of the independent variables for slope and intercept for each trace. In fitting this data set with six concentrations of substrate and then fitting the rate versus concentration to a hyperbola, a total of 14 independent parameters were used. It is better to fit the data globally to derive only the two independent parameters from the primary data using computer simulation based on numerical integration of the rate equations [9].

In fitting steady state data by simulation, we start with the full realistic model for an enzyme-catalyzed reaction including five rate constants and then make approximations to simplify the model to be consistent with the information content of the data and the desired steady state parameters.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \xrightarrow{k_3} E + P$$

One could fit the data using all five rate constants, then calculate the steady state kinetic parameters from Equation 2. It is well known that steady state kinetics cannot define intrinsic rate constants; a corollary of that statement is that a large combination of intrinsic rate constants can be found to fit the data and provide estimates of steady state kinetic parameters. Thus, any combination of rate constants that fit the data provides valid estimates for the steady state parameters (k_{cat}, K_m and k_{cat}/K_m). However, there will be large errors on each rate constant

because of the large number of combinations of rate constants that can account for the data. The large errors would then propagate to large error estimates for each steady state parameter, which would not provide a realistic estimate of the uncertainty in measuring each parameter.

To limit the number of variables, we lock three of the parameters at reasonable values to reduce the model to only two unknowns. For example, we can modify the model to mimic rapid equilibrium binding. To do so, we use a conservative estimate for diffusion-limited substrate binding (k_1 = 100 μM^{-1}s^{-1}) then make the chemistry step irreversible and product release at least 100-fold faster than k_2 so that the rate of chemistry defines k_{cat}. This represents the standard (often erroneous) interpretation of k_{cat} and K_m. However, because we are only using this approximation to fit steady state kinetic data, this model need not to be true physically to give valid estimates of the steady state kinetic parameters. The simplified model shown below gives estimates of k_{cat} and K_m directly.

$$E + S \underset{k_{-1}}{\overset{100\ \mu M^{-1}s^{-1}}{\rightleftharpoons}} ES \underset{0}{\overset{k_2}{\rightleftharpoons}} EP \xrightarrow{>> k_2} E + P$$
$$k_{cat} = k_2$$
$$K_m = k_{-1} / 100\ \mu M^{-1}s^{-1} \quad (4)$$

We can also use an alternative form of the model to obtain estimates of k_{cat} and k_{cat}/K_m directly. Here by setting k_{-1} = 0, after substrate binds it is always converted to product so k_{cat}/K_m is defined by the value of k_1. This model gives estimates of k_{cat} and k_{cat}/K_m from the global fit.

$$E + S \underset{0}{\overset{k_1}{\rightleftharpoons}} ES \underset{0}{\overset{k_2}{\rightleftharpoons}} EP \xrightarrow{>> k_2} E + P$$
$$k_{cat} = k_2$$
$$k_{cat} / K_m = k_1 \quad (5)$$

Again, it is important to note that this need not represent physical reality in defining the intrinsic rate constants; the approximations are acceptable because we are only using the results to compute the steady state kinetic parameters. In fact, we illustrate below that either model can be used to fit the data to give identical steady state parameters although the standard errors will differ.

In Figure 3, we show the results of fitting the same data used in Figure 2. In Figure 3A, the curves represent the global data fit using only two unknown parameters, k_{cat} and K_m (Equation 4) or k_{cat} and k_{cat}/K_m (Equation 5). Because the results of the two fitting methods are indistinguishable graphically, we only show one figure to represent both methods (Figure 3A). However, as summarized in Table 1, the error estimates vary depending on

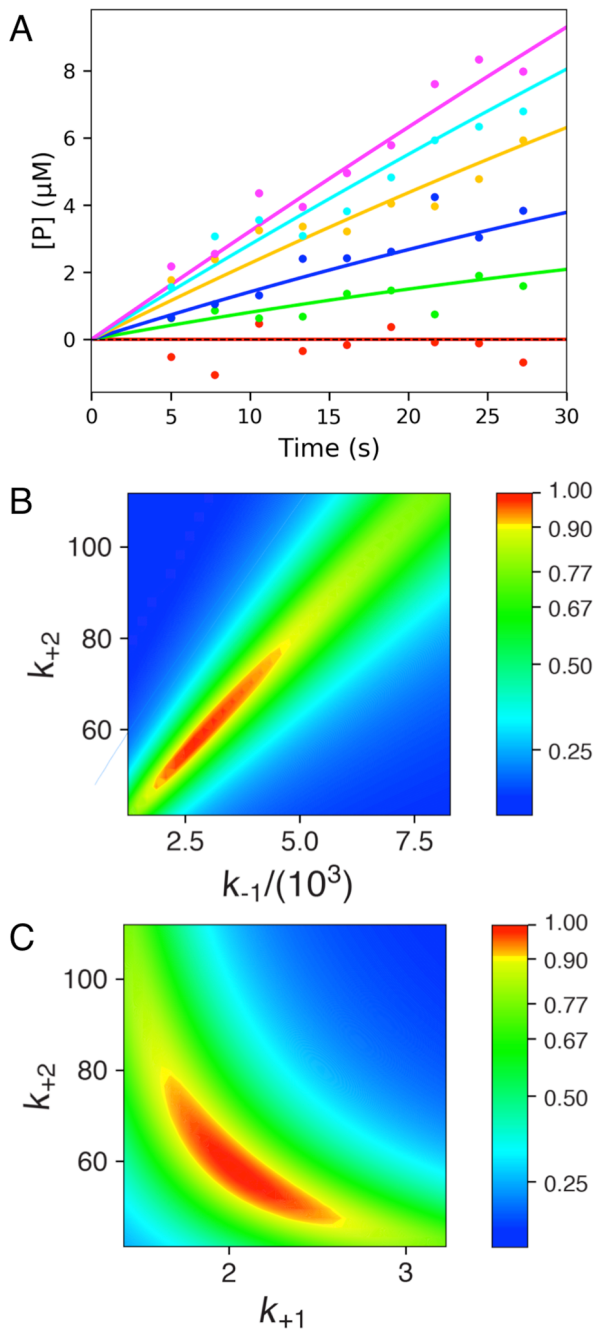


Figure 3: Fitting steady state data by simulation. A) Synthetic data from Figure 2A were fit globally to derive estimates of k_{cat} and K_m or k_{cat} and k_{cat}/K_m as described in the text. B) Confidence contour analysis from fitting to derive k_{cat} and K_m. C) Confidence contour analysis from fitting to derive k_{cat} and k_{cat}/K_m. The bar gives the color coding for the normalized χ^2 values [10].

the method used. As seen previously with equation-based data fitting, using the model to define k_{cat}/K_m directly is more precise than computing k_{cat}/K_m from individual estimates of k_{cat} and K_m. It should also be noted that either method of fitting data by simulation is more accurate than the corresponding equation-based data fitting. This is because we are fitting the entire data set using only two parameters rather than fourteen. Using extraneous parameters introduces additional errors in data fitting.

Fitting by simulation also affords confidence contour analysis to more precisely estimate errors in data fitting and to reveal relationships between parameters [10]. In Figure 3B we show the confidence contour analysis from fitting to derive $k_{cat} = k_2$ and $K_m = k_{-1}/k_1$. In this analysis, the two parameters are varied systematically and then we compute χ^2, which we plot using a color scale to represent z-axis values on the plot of k_2 versus k_{-1}. The colors represent the values of χ^2 normalized relative to the best fit so the red area defines the combinations of rate constants that give a good fit while the yellow band shows the χ^2 boundary surrounding a good fit [10]. The elongated zone of good fit illustrates the linear relationship between k_2 (defining k_{cat}) and k_{-1} (defining K_m). That is, this analysis clearly shows that the ratio of k_2/k_{-1} is known with greater certainty than either parameter alone. In this simplified model $k_{cat}/K_m = k_2k_1/k_{-1}$ (k_1 is fixed). Thus, the confidence contour analysis reveals that the data define k_{cat}/K_m more accurately than either k_{cat} or K_m.

The confidence contours for the global fit to derive k_1 (k_{cat}/K_m) and k_2 (k_{cat}) are shown in Figure 3C. The curvature of the red area fits an equation of the form $k_1 \times k_2$ = constant. This merely states that the net rate of product formation is a function of the combined rates of substrate binding and chemistry and that the net rate is known with greater certainty than either rate constant alone.

This analysis supports two important conclusions: (1) it is better to fit steady state data to define k_{cat} and k_{cat}/K_m rather than k_{cat} and K_m; and (2) simulation affords more accurate data fitting than the traditional methods of fitting to equations. Fitting data to equations necessarily involves limitations to conform to the approximations in defining the initial velocity of turnover before the substrate is consumed and product builds up, and it requires that the data be fit a second time in the form a plot of estimated rate versus concentration. Fitting by computer simulation overcomes these limitations.

Full progress curve analysis

The ability to fit data by simulation based on the numerical integration of rate equations frees the experimentalist from the confines of initial velocity measurements. One can easily follow the reaction to completion beyond the linear phase and even fit the entire time course to derive estimates of k_{cat} and k_{cat}/K_m. To illustrate this, we simulated ten data points at the same concentrations of substrate examined in Figure 2, but here we allow the reaction to go to completion (Figure 4A). The same standard deviation (0.5) now leads to less apparent noise because of the larger signal. The synthetic data were then globally fit to derive estimates of either k_{cat} and K_m (Equation 4) or k_{cat} and k_{cat}/K_m (Equation 5). Like before, the choice of method for data fitting did not affect the appearance of the fitted curves so we show only one (Figure 4A). However, the confidence contour analysis again shows the linear relationship between k_2 (defining k_{cat}) and k_{-1} (defining K_m), demonstrating that k_{cat}/K_m is more accurately defined by the data than either constant individually (Figure 4B). This analysis also revealed the lower standard errors estimated for k_{cat}/K_m measured directly compared to values calculated from the ratio of k_{cat} and K_m (see Table 1).

Analysis of full progress curve kinetics provides the most accurate estimates. Given the experimental constraints of limited substrate concentrations and the same number of data points collected, it is better to spread the data points out and follow the reaction to completion rather than restrict the measurement to the initial velocity. One could stop data collection at any time and still be able to fit the data without regard for maintaining initial velocity conditions. Fitting data as was done by Michaelis and Menten more than 100 years ago imposes significant limitations on the quality of data that can be collected because it restricts data fitting to the earliest part of the curve with the only small amounts of product formed. It is more accurate to monitor the reaction for longer times, allow the reaction rate to fall off as substrate is consumed but account for the deviation from linearity by fitting the data using computer simulation. Product inhibition can also be resolved if it contributes significantly to the data [11].

Standards for data collection

When it is possible, data should be collected at substrate concentrations exceeding the K_m by at least 4-fold so that the data reach 80% saturation. Concentrations 9-fold greater than the K_m are required to reach 90% saturation. The question of how high to go in substrate concentration must also be considered relative to the availability and solubility of the substrate. The standard rules apply for measurement of initial velocities requiring that less than 10% of the substrate should be consumed during the measurement to support the steady state approximation. Of course, this requirement does not apply if the data are fit by computer simulation, so more accurate data can be obtained based on formation of a larger signal in measuring product formation.

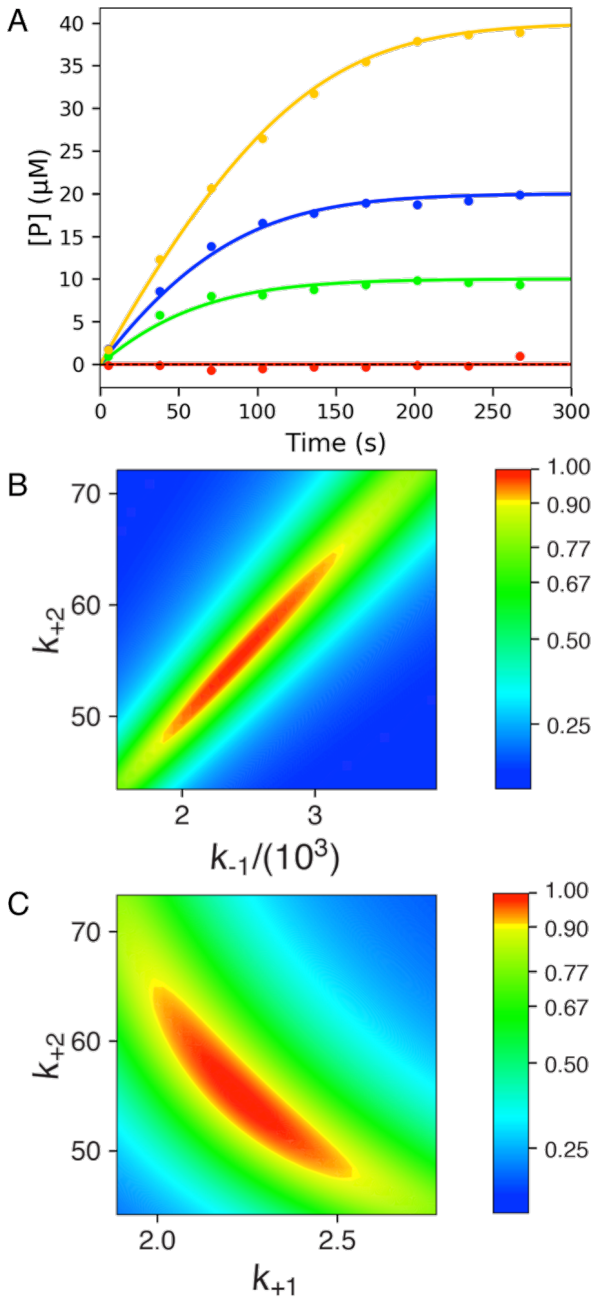


Figure 4: Fitting full progress curve data by simulation. A) Synthetic data were generated and then were fit globally to derive estimates of k_{cat} and K_m or k_{cat} and k_{cat}/K_m as described in the text. B) Confidence contour analysis from fitting to derive k_{cat} and K_m. C) Confidence contour analysis from fitting to derive k_{cat} and k_{cat}/K_m. The bar gives the color coding for the normalized χ^2 values [10].

It is always important to carefully select the minimal number of measurements to provide the needed information to optimally use limited resources. Here, full time course measurements are by far the best, as described above. In the absence of product inhibition, steady state kinetic parameters can be derived from a single sample starting at a substrate concentration 8–10 fold higher than the K_m and following the reaction to completion. To test for product inhibition, two replicates at lower substrate concentrations will suffice. At larger substrate concentrations, larger concentrations of product formed toward the end of the reaction alter the rate of approach to equilibrium if the rebinding of product to the enzyme occurs appreciably. Globally fitting measurements at three substrate concentrations may be sufficient to define k_{cat} and k_{cat}/K_m and K_i for product inhibition. The ready availability of computer programs for fitting by numerical integration of the rate equations renders the initial velocity measurements obsolete.

In setting up initial velocity measurements one must first decide on the range of concentrations to use and whether to space the points evenly. It is generally accepted that concentrations should be more closely spaced below and slightly above the K_m and spaced further apart at higher concentrations. The data at low concentration define k_{cat}/K_m while the data at the higher concentrations are only needed to extrapolate to get k_{cat}. Cleland has suggested collecting data with points evenly spaced on a Lineweaver–Burk plot [12]. However, this conclusion represents a mistake rooted in the distortion of the data when viewed on a double reciprocal plot as shown in Figure 5A. Spacing points evenly on a double reciprocal plot does not provide the best distribution of data given the importance of accurately defining k_{cat} and k_{cat}/K_m. A better alternative is to space points evenly on a logarithmic scale (Figure 5B,C). Here 11 points can be distributed on a log scale with [S]/K_m ratios of 0.16, 0.25, 0.4, 0.6, 1, 1.6, 2.5, 4, 6 and 10 (rounded off). This provides a convenient series that most accurately defines both k_{cat} and k_{cat}/K_m. These guidelines are predicated on having an estimate of k_{cat} and k_{cat}/K_m before setting up the measurements. Since all experiments must be replicated prior to publication, the first experiment can be used to explore the range of concentrations and time for data collection. A second experiment can then be designed based upon the initial estimates to achieve an optimal distribution of data points to get publication quality data.

Another question is whether to collect triplicates at each concentration or to sample at three times as many concentrations. Because all independent measurements are treated equally in fitting by nonlinear regression, it is better to collect data at more concentrations rather than in triplicate at fewer concentrations. The average standard deviation of the measurements is evident in the scatter of the data from the fitted curve and can be estimated from the χ^2 valued derived in fitting the data.

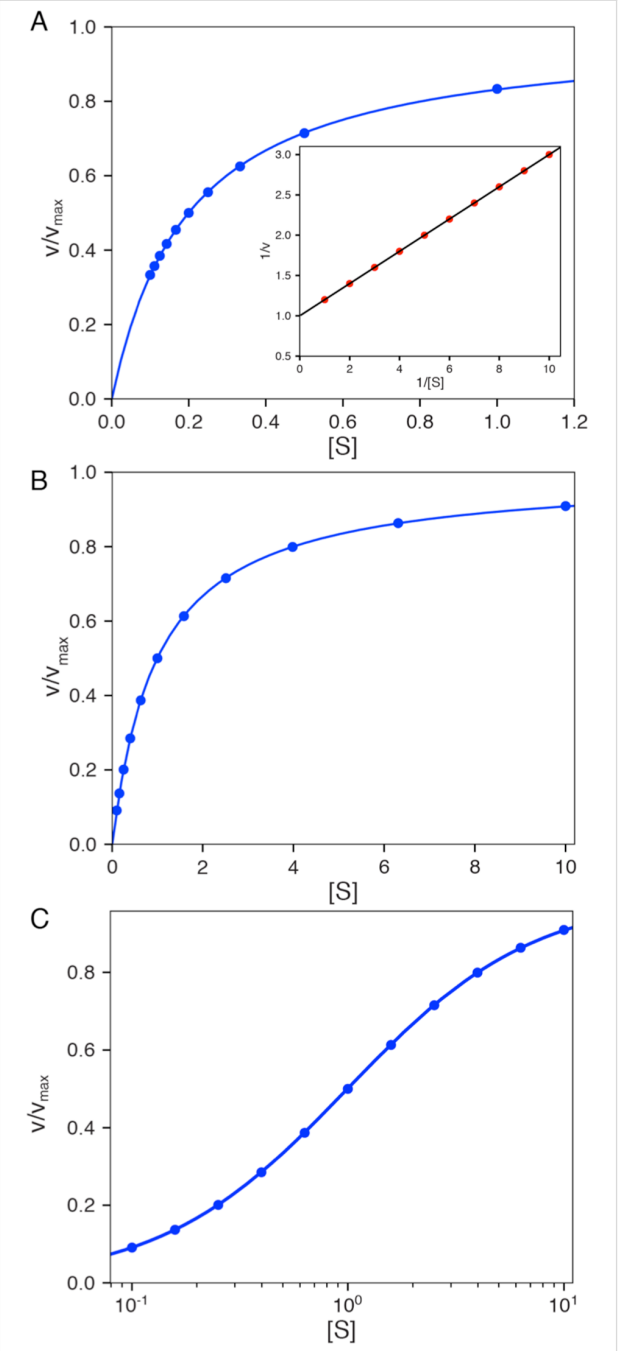


Figure 5: Optimal spacing of data points. Different scenarios for computing the distribution of data points for steady state measurements are shown. A) Linear distribution on a double reciprocal plot. Sample data were calculated with k_{cat} = 1 s^{-1} and K_m = 0.2 (arbitrary units). The inset shows the same data on a double reciprocal plot. B) Logarithmic distribution of data points. Sample data were calculated with k_{cat} = 1 s^{-1} and K_m = 1 (arbitrary units). C) Data in B on a logarithmic *x*-axis. The smooth lines show the fitted curve.

Active site concentration

Interpretation of steady state turnover rates is dependent on an accurate estimate of the concentration of active sites. There are significant systematic errors in measurements of protein concentrations using dye-binding assays or by absorbance measurements, and the fraction of protein that is active is not known without direct measurement. For these reasons, it is important to perform an active site titration to establish the concentration of active sites. One method is isothermal titration calorimetry relying on the heat change upon binding of a substrate analog. Because the method is relatively insensitive it requires high concentrations of protein (usually μM) so the stoichiometry is easy to determine when titrating with a known concentration of a substrate analog [13]. In addition, many proteins show a change in fluorescence (tyrosine and tryptophan residues) upon substrate binding, affording accurate measurements of the stoichiometry and dissociation constant for binding from an equilibrium titration [14]. Other methods, such as rapid gel filtration and filter binding assays are limited by the rate of ligand dissociation relative to the time required to perform the separation. Alternatively, the kinetics of a pre-steady state burst of product formation can be used to estimate the concentration of active sites under favorable conditions [15]. In any event, kinetic data should be normalized by dividing the rate by the concentration of enzyme active sites, and the basis for estimating enzyme concentration should be clearly stated. It is no longer acceptable to report enzyme specific activity in units of product/min/mg of enzyme; rather, report values of k_{cat} and k_{cat}/K_m.

Interpretation of k_{cat}/K_m

The steady state kinetic parameter, k_{cat}/K_m is not merely the ratio of k_{cat} and K_m; rather, it should be considered as a single parameter because it quantifies enzyme specificity, efficiency and proficiency [4,5]. Intuitively, it may seem that the substrate with the greater k_{cat} reacts faster and would be preferred, but that is not necessarily the case when two competing substrates are present as the one with a lower K_m would occupy more of the enzyme. However, simple algebra shows that the relative rate of turnover of two competing substrates is defined by their relative concentrations and k_{cat}/K_m values for substrates A and B.

$$\frac{v_A}{v_B} = \frac{\left(k_{cat,A} / K_A\right)[A]}{\left(k_{cat,B} / K_B\right)[B]}$$

Thus, enzyme specificity is quantified by k_{cat}/K_m. It is for this reason that k_{cat}/K_m is called the specificity constant. Specificity is a function of the apparent second order rate constants for substrate binding and conversion to product. When competing substrates are both present, the one that binds to the enzyme the fastest and is then converted to product wins the competition. In contrast, k_{cat} defines how fast the enzyme catalyzes a given reaction, not which substrate the enzyme prefers.

Although k_{cat}/K_m provides a quantitative measure of enzyme specificity, it does not define the underlying basis for specificity. Therefore, a major effort is currently underway to understand how elementary steps in catalysis contribute to the observed specificity and to understand how enzymes evolve to acquire new specificities. Analysis of enzyme families has revealed that members within a family share a common catalytic core and a variable loop domain that closes over the substrate and confers substrate specificity [16,17]. Moreover, specificity could be dependent on conformational changes in the loop domain after substrate binding. The concept of induced-fit, where the substrate induces a change in enzyme structure to align catalytic residues, was first proposed in somewhat vague terms by Koshland [18] in an attempt to understand how an enzyme can exclude a smaller substrate than the preferred one.

The induced-fit model proposes a two-step binding pathway in which the substrate first binds to an open form of the enzyme and then the enzyme closes leading to tighter binding and organization of catalytic residues.

$$\overset{open}{E} + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \overset{open}{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \overset{closed}{FS} \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} \overset{closed}{FP} \xrightarrow{k_4} \overset{open}{E} + P$$

For decades debate raged as to whether a two-step binding mechanism could lead to increased enzyme specificity. Most notably, Fersht argued that because a two-step binding sequence has the same net free energy change as a corresponding one-step mechanism, a two-step binding sequence could not lead to greater enzyme specificity [5]. This logic is flawed because it follows from the simple definition of K_m as equal to the K_d for substrate binding and assumes the conformational change step is fast and at equilibrium preceding catalysis. Thus, the conclusion is a restatement of the assumptions used to define the model.

More recently, Warshel has asserted that pre-chemistry barriers cannot contribute to enzyme specificity unless they are rate limiting [19]. In his arguments, Warshel fails to appreciate the distinction between steps in the pathway that govern specificity (k_{cat}/K_m) versus those that govern the net turnover rate (k_{cat}). The terminology in which the specificity constant is given by the ratio of k_{cat} divided by K_m contributed to the confusion. Throughout his recent paper, Warshel continually referred to the rate-limiting step as if it also defined specificity. In general, it does not.

To resolve this controversy, a direct measurement of the rates of the conformational change and the chemical reaction at the active site of the enzyme was required. Steady state kinetic methods do not suffice. Transient state kinetic analysis are needed to measure events occurring during a single enzyme cycle, but in the end, we must account for steady state kinetic parameters calculated from intrinsic rate constants. Resolution of the longstanding controversy over the role of induced-fit in enzyme specificity illustrates the importance of properly interpreting k_{cat} and k_{cat}/K_m based on asking how each step in the reaction contributes to the observed k_{cat}/K_m values for the correct and incorrect substrates.

Figure 6 shows three possible scenarios for the effect of the conformational change on k_{cat} and k_{cat}/K_m. In this figure, we show free energy profiles computed from different combinations of rate constants for a minimal three-step reaction where product release is fast after the chemistry step.

$$\overset{open}{E} + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \overset{open}{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \overset{closed}{FS} \xrightarrow{k_3} \overset{open}{E} + P$$

$$k_{cat}/K_m = k_1 \frac{k_2 k_3}{k_2 k_3 + k_{-1}(k_{-2} + k_3)}$$

In each figure, the slow step in the pathway defines k_{cat} and is identified as the step with the largest local barrier (relative to the local minimum) in the free energy profile. On the other hand, the specificity constant, k_{cat}/K_m, can be identified as the steps leading from the starting state up to the highest overall barrier.

Case 1

In Figure 6A, k_{cat}/K_m and k_{cat} are both governed by the chemistry step. In this model, the initial binding and conformational change are both rapid equilibrium reactions leading up to chemistry. In this case, k_{cat} and k_{cat}/K_m can be approximated as follows (Note $K_i = k_i/k_{-i}$):

$$\text{If } k_{-1},\ k_2,\ k_{-2} >> k_3$$

$$k_{cat} = \frac{K_2 k_3}{1 + K_2}$$

$$k_{cat}/K_m = K_1 K_2 k_3$$

Note that k_{cat} is not simply defined by k_3; rather, the equilibrium constant for the conformational change step defines the fraction of the bound substrate that is in the *FS* state ($K_2/(1 + K_2)$). An unfavorable equilibrium constant for the conformational change (K_2) could reduce both k_{cat} and k_{cat}/K_m.

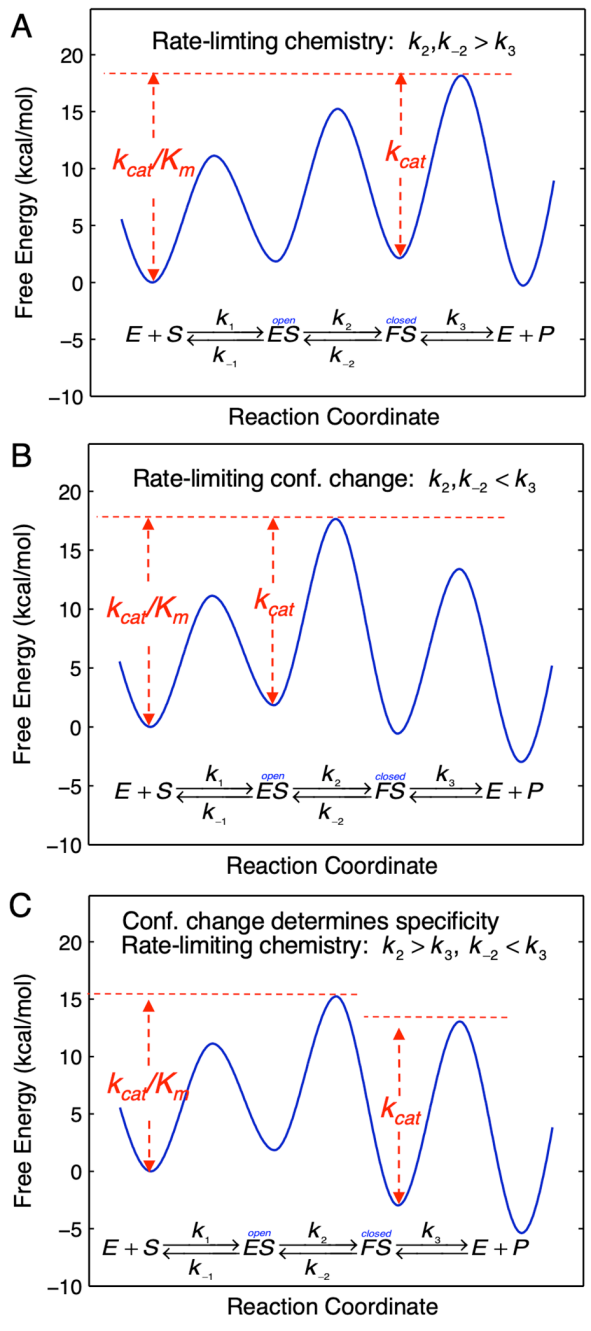


Figure 6: Free energy profiles. Free energy profiles are shown for a three step model with different rate constant relationships as described in the text and summarized on each figure. The free energy profile was calculated using transition state theory: $\Delta G^{\ddagger} = -RT \cdot \ln(k/(k_B T/h))$, where k is the rate constant, k_B is the Boltzmann constant and h is Planck's constant. Second order rate constants were converted to pseudo-first order rate constants using an estimated physiological concentration of substrate.

Case 2

We next consider the case shown in Figure 6B where the conformational change is rate-limiting. Here it can be seen that the rate of the conformational change governs both k_{cat} and k_{cat}/K_m.

$$\text{If } k_2,\ k_{-2} << k_3$$

$$k_{cat} = k_2$$

$$k_{cat} / K_m = K_1 k_2$$

This model mimics the standard view of catalysis with a single equilibrium binding step followed by a single rate limiting step, but in this case, the conformational change, not chemistry, is rate limiting.

Case 3

Finally, we consider the case where chemistry is rate-limiting, but the reverse of the conformational change step is slower than the rate of chemistry (Figure 6C). Here we see that the conformational change step governs specificity (k_{cat}/K_m) but the rate of chemistry governs k_{cat}.

$$\text{If } k_2 >> k_3 \text{ and } k_{-2} << k_3$$

$$k_{cat} = k_3$$

$$k_{cat} / K_m = K_1 k_2$$

$$K_m = k_3 / K_1 k_2$$

This leads to a surprising result that had not been anticipated in decades of research. To fully understand this, it is instructive to examine the equation for k_{cat}/K_m calculated from the three-step model (Equation 2).

$$k_{cat} / K_m = \frac{k_1 k_2 k_3}{k_2 k_3 + k_{-1}(k_{-2} + k_3)}$$

When $k_{-2} << k_3$, this reduces to:

$$k_{cat} / K_m = \frac{k_1 k_2 k_3}{k_2 k_3 + k_{-1} k_3}$$

By dividing the numerator and denominator by k_3, this reduces to an equation that no longer includes k_{cat} (k_3).

$$k_{cat} / K_m = \frac{k_1 k_2}{k_2 + k_{-1}}$$

This equation can be further reduced by assuming that the substrate binding to the open state is in rapid equilibrium, i.e., $k_{-1} >> k_2$.

$$k_{cat} / K_m = K_1 k_2$$
$$k_{cat} = k_3$$

This leads to the surprising result that the K_m is defined by:

$$K_m = \frac{k_{cat}}{k_{cat} / K_m} = \frac{k_3}{K_1 k_2}$$

The product, K_1k_2 defines the second order rate constant for substrate binding. Thus the K_m is defined the balance between the rate of enzyme turnover relative to the rate of substrate binding. Because the reverse of the conformational change step is very slow, the two-step binding reaction does not come to equilibrium. Rather, the substrate binds and the enzyme closes leading to rapid catalysis and product release. Because the reverse of the conformational change step is so much slower than chemistry, the initial weak substrate binding and the conformational change are the primary determinants of specificity.

DNA polymerase fidelity

DNA polymerases provide ideal model systems to study enzyme specificity because fidelity is high and physiologically relevant, and the alternate substrates are well known. Moreover, it is easy to perform single turnover kinetic measurements to examine steps leading up to the chemical reaction by mixing an enzyme DNA complex with only one nucleoside triphosphate. Recent work on DNA polymerase fidelity has shown that the rate of the conformational change from open to closed state is much faster than chemistry [20,21]. If we were only concerned with defining the rate-limiting step (k_{cat}) we would stop at this point and simply conclude that chemistry was rate limiting; and since k_{cat}/K_m defines specificity, the chemistry step must also define specificity. However, that would be wrong. An additional experiment was required to measure the rate of substrate release using dideoxy-terminated DNA to allow the conformational change but prevent chemistry. This experiment allowed the measurement of the rate of enzyme reopening to release substrate before chemistry. The results showed that once the enzyme closes over a correct substrate, it almost always continues to react rather than release the bound substrate. Globally fitting multiple experiments yielded the following rate constants [21]:

$$\underset{}{\overset{open}{ED_n + N}} \xrightarrow{200\ \mu M} \underset{}{\overset{open}{ED_n N}} \underset{1\ s^{-1}}{\overset{2000\ s^{-1}}{\rightleftharpoons}} \underset{}{\overset{closed}{FD_n N}}$$

$$\xrightarrow[chemistry]{20\ s^{-1}} \underset{}{\overset{closed}{FD_{n+1} PP_i}} \xrightarrow{>> k_3} \underset{}{\overset{open}{ED_{n+1} + PP_i}}$$

where ED_n represents an enzyme–DNA complex in the *open* state with a primer strand *n* residues long, while FD_nN represents the closed state with nucleotide (*N*) bound. Note that we show $1/K_1$ = 200 μM for the initial weak binding step.

The initial weak binding of nucleotide to the open state (K_d = 200 μM) is followed by a very fast conformational change to the closed state to afford a net $K_d = 1/(K_1(1 + K_2))$ = 200 nM. Because the chemistry step (k_3) is so much faster than the rate at which the enzyme opens to release the substrate (k_{-2}), the FD_nN state goes forward 95% of the time. Thus, the conformational change is the primary determinant of enzyme specificity because it commits the substrate to forward reaction. For the DNA polymerase studied, the rate of product release is much faster than chemistry so the model reduces to a three-step model. Accordingly the specificity constant is defined by the two-step binding reaction, while k_{cat} is defined by the rate of the chemical reaction.

This result was a big surprise, which had not been anticipated in attempts to foresee the ways in which induced-fit could contribute to specificity [22]. For 20 years numerous investigators in the DNA polymerase field had attempted to resolve whether the conformational change or chemistry was rate limiting. We had neglected to measure the rate of the reverse of the conformational change (enzyme opening to allow release of bound substrate) relative to the rate of chemistry, and that proved to be the key to understanding specificity. As shown in Figure 7, the free energy profile shows that after the conformational change, the enzyme is committed to go forward because there is a larger barrier to going backwards. The highest overall barrier is the conformational change step, thus defining the specificity constant [21,23,24].

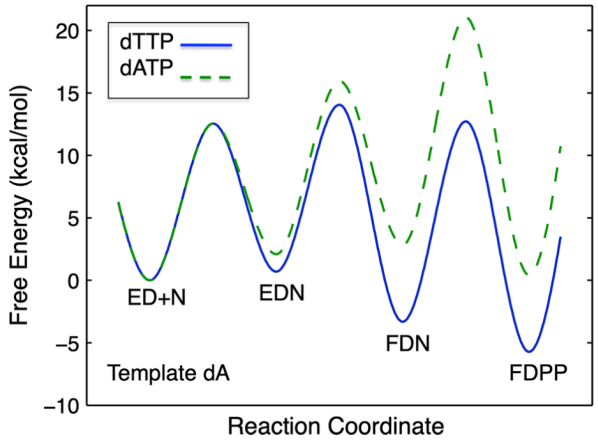


Figure 7: Free energy profile for DNA polymerization. Free energy profiles for a correct base pair (solid blue line) and a mismatch (dashed green line) were computed from data presented in [21].

$$k_{cat} / K_m = K_1 k_2 = 10\ \mu M^{-1} s^{-1}$$

$$k_{cat} = k_3 = 20\ s^{-1}$$

$$K_m = k_3 / K_1 k_2 = 2\ \mu M$$

$$K_d = \frac{1}{K_1(1+K_2)} = 0.1\ \mu M$$

We compare a free energy profile for correct nucleotide incorporation with that for a mismatch (Figure 7). With a mismatch (dashed line), the chemistry step becomes very slow, while the rate of enzyme opening is much faster. Thus, for a mismatch, the conformational change step comes to equilibrium prior to rate-limiting chemistry. In this case, the chemistry step governs both specificity and rate-limiting steps.

$$k_{cat} = \frac{K_2 k_3}{1+K_2}$$

$$k_{cat} / K_m = K_1 K_2 k_3$$

Mismatch recognition by the enzyme leads to a change in the specificity-determining step, but not the rate-limiting step.

We can now understand that the conformational change is the major specificity-determining step. The substrate binds weakly and then the enzyme closes. If the substrate shows the right geometry (structurally and electrostatically) the closed state is stabilized and organization of catalytic residues leads to fast catalysis. If the substrate is not the right size and shape, the enzyme fails to close tightly, chemistry is slow, and the enzyme rapidly opens to release the mismatched substrate [23,24].

This new paradigm for enzyme specificity provides a very satisfying resolution of the long-standing controversy over the role of induced fit in enzyme specificity. The conformation change serves as a gate-keeper to facilitate catalysis of the favored substrate while promoting release of alternate substrates.

Conclusion

This short review shows that the traditional Michaelis–Menten equation defined in terms of k_{cat} and K_m should be replaced by one in which the two variable parameters are k_{cat} and k_{cat}/K_m. There are two reasons for this change: (1) k_{cat}/K_m is the most important steady state kinetic parameter because it quantifies enzyme specificity, efficiency and proficiency; and (2) there are smaller errors in fitting to derive k_{cat}/K_m directly rather than by calculation of the ratio of k_{cat} and K_m derived independently in fitting steady state kinetic data. In addition, there are significant advantages in fitting by computer simulation rather than in using the conventional approach using equations. Instead of fitting steady state data to a straight and then fitting the concentration dependence of the observed rate, the raw data can be fit directly in a single step with fewer unknown variables, resulting in less error on the estimates for steady state kinetic parameters.

The use of the ratio k_{cat}/K_m to describe the specificity constant has long been a source of confusion. We now recognized that k_{cat}/K_m and k_{cat} can reflect different steps in the enzyme pathway. Although k_{cat} is a function of rate limiting steps in the pathway, steps defining k_{cat}/K_m establish specificity and need not be identical to the rate-limiting steps. Here the longstanding use of k_{cat}/K_m as the specificity constant gets in the way of proper understanding because, of course, one expects that k_{cat} is part of k_{cat}/K_m so they must be measuring the same step. This simplified logic overlooks the situation where k_{cat} in both the numerator and denominator of k_{cat}/K_m cancel so that the ratio is no longer related to k_{cat}; such is the case for DNA polymerase specificity.

Results presented here also document the advantages of fitting kinetic data using computer simulation based on the numerical integration of rate equations. Beyond what is shown here, one can also simultaneously fit steady state data along with equilibrium binding and pre-steady data kinetic data to derive a single unifying model to account for all of the results. This approach provides the most robust and accurate method for data fitting to ensure that the model fully accounts for all experimental observations. Moreover, confidence contour analysis provides a critical check to show the extent to which the fitted parameters are constrained by the data, and thereby support the model.

Financial Conflict of Interest Statement

KAJ is President of *KinTek Corporation* which licenses the *KinTek Explorer* software described in this review.

Acknowledgements

This work was supported by the Welch Foundation (F-1604).

ORCID® iDs

Kenneth A. Johnson - https://orcid.org/0000-0002-6575-2823

References

1. Michaelis, L.; Menten, M. L. *Biochem. Z.* **1913,** *49,* 333–369.
2. Johnson, K. A.; Goody, R. S. *Biochemistry* **2011,** *50,* 8264–8269. doi:10.1021/bi201284u
3. Briggs, G. E.; Haldane, J. B. S. *Biochem. J.* **1925,** *19,* 338–339. doi:10.1042/bj0190338
4. Miller, B. G.; Wolfenden, R. *Annu. Rev. Biochem.* **2002,** *71,* 847–885. doi:10.1146/annurev.biochem.71.110601.135446

5. Fersht, A. R. *Enzyme Structure and Mechanism,* 3rd ed.; Freeman: New York, 1999.
6. Cleland, W. W. *Steady State Kinetics*. In *The Enzymes;* Boyer, P. D., Ed.; Elsevier, 1970; pp 1–65. doi:10.1016/s1874-6047(08)60180-8
7. Lineweaver, H.; Burk, D. *J. Am. Chem. Soc.* **1934,** *56,* 658–666. doi:10.1021/ja01318a036
8. Cleland, W. W. *Nature* **1963,** *198,* 463–465. doi:10.1038/198463a0
9. Johnson, K. A.; Simpson, Z. B.; Blom, T. *Anal. Biochem.* **2009,** *387,* 20–29. doi:10.1016/j.ab.2008.12.024
10. Johnson, K. A.; Simpson, Z. B.; Blom, T. *Anal. Biochem.* **2009,** *387,* 30–41. doi:10.1016/j.ab.2008.12.025
11. Johnson, K. A. *Methods Enzymol.* **2009,** *467,* 601–626. doi:10.1016/s0076-6879(09)67023-3
12. Cleland, W. W. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1967,** *29,* 1–32.
13. Qian, Y.; Johnson, K. A. *J. Biol. Chem.* **2017,** *292,* 13068–13084. doi:10.1074/jbc.m117.791392
14. Anderson, K. S.; Sikorski, J. A.; Johnson, K. A. *Biochemistry* **1988,** *27,* 1604–1610. doi:10.1021/bi00405a032
15. Johnson, K. A. *Methods Enzymol.* **1995,** *249,* 38–61. doi:10.1016/0076-6879(95)49030-2
16. Lahiri, S. D.; Zhang, G.; Dai, J.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2004,** *43,* 2812–2820. doi:10.1021/bi0356810
17. Zhang, G.; Mazurkie, A. S.; Dunaway-Mariano, D.; Allen, K. N. *Biochemistry* **2002,** *41,* 13370–13377. doi:10.1021/bi026388n
18. Koshland, D. E. In *The Enzymes;* Boyer, P. D.; Lardy, H.; Myrback, K., Eds.; Academic: New York, 1959; pp 305–346.
19. Ram Prasad, B.; Kamerlin, S. C. L.; Florián, J.; Warshel, A. *Theor. Chem. Acc.* **2012,** *131,* 1288. doi:10.1007/s00214-012-1288-6
20. Tsai, Y.-C.; Johnson, K. A. *Biochemistry* **2006,** *45,* 9675–9687. doi:10.1021/bi060993z
21. Kellinger, M. W.; Johnson, K. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 7734–7739. doi:10.1073/pnas.0913946107
22. Herschlag, D. *Bioorg. Chem.* **1988,** *16,* 62–96. doi:10.1016/0045-2068(88)90038-7
23. Kirmizialtin, S.; Nguyen, V.; Johnson, K. A.; Elber, R. *Structure* **2012,** *20,* 618–627. doi:10.1016/j.str.2012.02.018
24. Kirmizialtin, S.; Johnson, K. A.; Elber, R. *J. Phys. Chem. B* **2015,** *119,* 11513–11526. doi:10.1021/acs.jpcb.5b05467

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Computational characterization of enzyme-bound thiamin diphosphate reveals a surprisingly stable tricyclic state: implications for catalysis

Ferran Planas[1], Michael J. McLeish[*2] and Fahmi Himo[*1]

Full Research Paper

Open Access

Address:
[1]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, SE-10691 Stockholm, Sweden and [2]Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, Indianapolis IN 46202, USA

Email:
Michael J. McLeish[*] - mcleish@iupui.edu; Fahmi Himo[*] - fahmi.himo@su.se

* Corresponding author

Keywords:
binding site; DFT; enzyme mechanism; quantum chemical calculations; ThDP-dependent

Beilstein J. Org. Chem. **2019,** *15,* 145–159.
doi:10.3762/bjoc.15.15

Received: 26 September 2018
Accepted: 10 December 2018
Published: 16 January 2019

This article is part of the thematic issue "Enzymes in chemical transformations".

Guest Editor: K. N. Allen

Abstract

Thiamin diphosphate (ThDP)-dependent enzymes constitute a large class of enzymes that catalyze a diverse range of reactions. Many are involved in stereospecific carbon–carbon bond formation and, consequently, have found increasing interest and utility as chiral catalysts in various biocatalytic applications. All ThDP-catalyzed reactions require the reaction of the ThDP ylide (the activated state of the cofactor) with the substrate. Given that the cofactor can adopt up to seven states on an enzyme, identifying the factors affecting the stability of the pre-reactant states is important for the overall understanding of the kinetics and mechanism of the individual reactions.

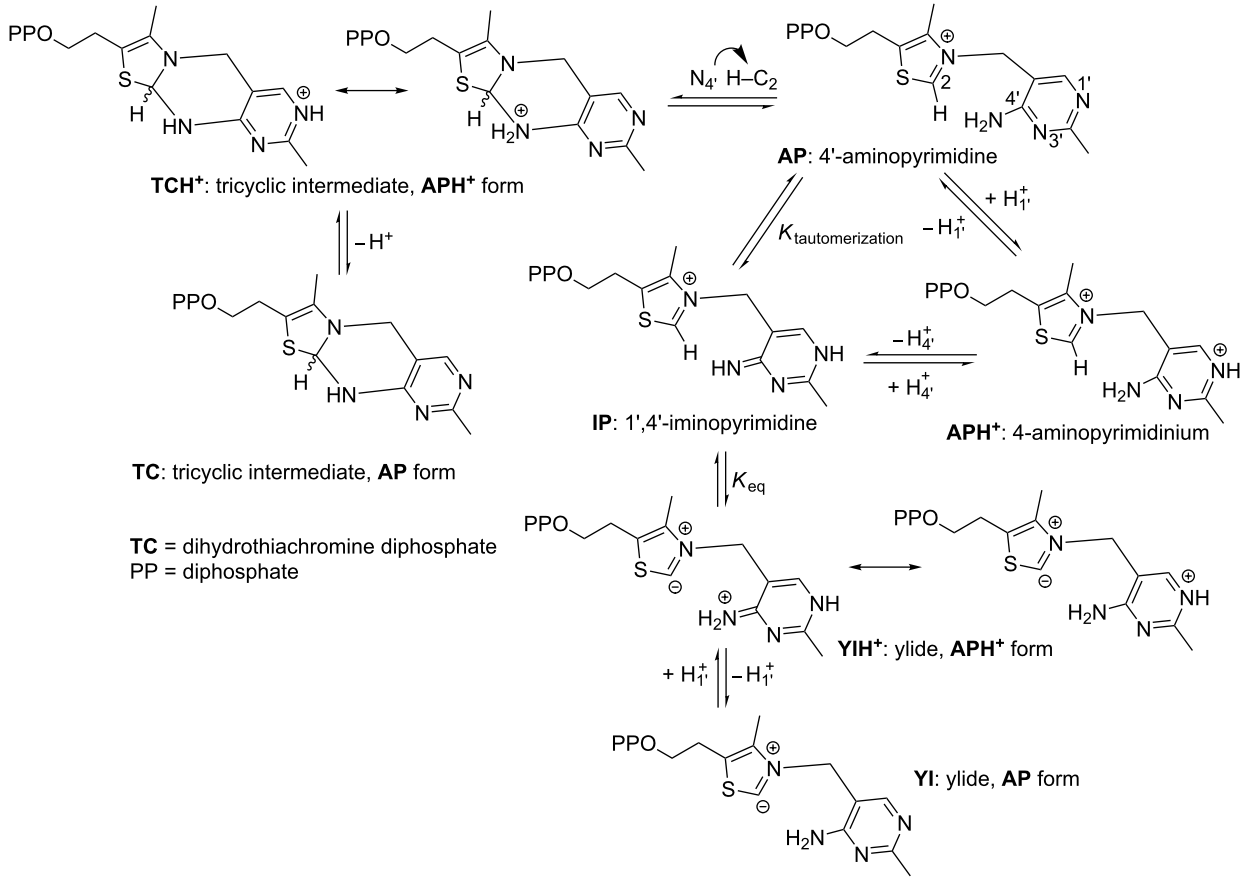
In this paper we use density functional theory calculations to systematically study the different cofactor states in terms of energies and geometries. Benzoylformate decarboxylase (BFDC), which is a well characterized chiral catalyst, serves as the prototypical ThDP-dependent enzyme. A model of the active site was constructed on the basis of available crystal structures, and the cofactor states were characterized in the presence of three different ligands (crystallographic water, benzoylformate as substrate, and (*R*)-mandelate as inhibitor). Overall, the calculations reveal that the relative stabilities of the cofactor states are greatly affected by the presence and identity of the bound ligands. A surprising finding is that benzoylformate binding, while favoring ylide formation, provided even greater stabilization to a catalytically inactive tricyclic state. Conversely, the inhibitor binding greatly destabilized the ylide formation. Together, these observations have significant implications for the reaction kinetics of the ThDP-dependent enzymes, and, potentially, for the use of unnatural substrates in such reactions.

Introduction

Enzymes that depend on thiamin diphosphate (ThDP, Scheme 1) can be found in a wide range of metabolic pathways. Although they are known to catalyze the formation of C–N, C–O and C–S bonds, ThDP-dependent enzymes generally catalyze the breakdown and formation of C–C bonds adjacent to a carbonyl group [1,2]. The resultant 2-hydroxyketones are often chiral, so these enzymes are being increasingly studied for their use as biocatalysts in the preparation of pharmaceuticals and agrochemicals [3]. ThDP is an unusual cofactor in that, even without the enzyme, it can catalyze many of these reactions [2]. For example, the decarboxylation of pyruvate in water can be accomplished by ThDP, but when it is bound to the enzyme pyruvate decarboxylase (PDC), the decarboxylation rate is increased by 12 orders of magnitude [2,4]. Clearly, the catalytic power of the cofactor is greatly enhanced by the enzyme-bound environment. A fundamental understanding of how this enhancement is achieved could potentially lead to the development of new and improved biocatalysts.

At a minimum, ThDP-catalyzed reactions all require the formation of a C2-carbanion or ylide [5] (Scheme 1). This is achieved through a series of proton transfers during which several different states of the cofactor are formed [6]. Starting from the neutral form of ThDP (**AP**), the cofactor can be protonated at the N1′ position, resulting in the **APH**$^+$ state. With only one known exception [7], the protonation/deprotonation of the N1′ position is performed by a highly conserved glutamic acid residue that is thought to stabilize the imino tautomer **IP** [6]. The subsequent loss of a proton from N4′ of **APH**$^+$ gives the **IP** state. Deprotonation of the C2 position results in the ylide form which can be either protonated (**YIH**$^+$) or deprotonated (**YI**) at the N1′ position. The C2 deprotonation is believed to be performed by the N4′ nitrogen [2,8,9], and is assisted by the cofactor being held in a "V" conformation in which the imino group is located within a hydrogen bonding distance of the C2 of the thiazolium ring [9-14].

While the importance of the catalytically critical ylide was readily recognized, obtaining evidence for the participation of the 4′-amino group and the imino tautomer **IP** proved more challenging. Initially, model compounds were used to identify the signature UV absorbances for the **IP** form of ThDP. These



Scheme 1: The variety of forms of enzyme-bound ThDP.

were then used to demonstrate the presence of **IP** on yeast PDC [15]. Subsequently, the **IP** form was shown to have a positive CD signal around 300–310 nm, while a negative peak around 320–330 nm, similar to that observed upon binding of ThDP to apo transketolase [16], was assigned to the **AP** form [15]. These, along with signature CD and UV signals for intermediates further along the reaction pathway, have now been observed for more than 10 ThDP-dependent enzymes [6,17]. As yet, no electronic signature has been observed for the $\mathbf{APH}^+$ form. However, solid-state NMR using ^{15}N and ^{13}C-labeled ThDP has been used to identify $\mathbf{APH}^+$ on pyruvate decarboxylase and the E1 component of the pyruvate dehydrogenase complex [18].

In addition to the plethora of experimental investigations, a number of computational studies have addressed issues regarding the various states of ThDP. For example, in some very early work, Jordan used semi-empirical methods to study the electronic structure and conformational space of the cofactor in the gas phase, acknowledging the difficulty of comparing these results to reactions in solution and on the enzyme [19,20]. Thirty years later, density functional theory (DFT) calculations showed that the 4′-amino moiety of the cofactor can either accept or donate a proton in the reactions, depending on the protonation state of N1′ [21].

Orbital analysis of the $\mathbf{IP}/\mathbf{YIH}^+$ reaction showed that full formation of ylide was dependent on deprotonation of N1′ and, consistent with experimental findings, deprotonation was, in turn, likely dependent on conformational changes induced by the presence of substrate [22]. More recently, the relative stabilities of a number of the ThDP states (**AP**, $\mathbf{APH}^+$, **IP** and **YI**) were obtained using DFT methods, employing a model of the cofactor along with the hydrogen-bonding carboxylate moiety [23]. Subsequently, a similar approach was used to characterize the nucleophilicity of the N1′ and N4′ centers [24]. In many cases, rather than simply focus on the cofactor, computational studies have been used to investigate full reaction mechanisms of ThDP enzymes, including pyruvate decarboxylase (PDC) [25-28], benzoylformate decarboxylase (BFDC) [29,30], acetohydroxy acid synthase [24,31-35], pyruvate dehydrogenase (PDH) [36], benzaldehyde lyase [37], cyclohexane dione hydrolase [38], oxalyl-CoA decarboxylase [39], DXP synthase [40] and transketolase [41,42].

It is surprising that almost none of these studies acknowledged that there is a second, albeit less well discussed, path for the ThDP cofactor, i.e., the formation of a tricyclic, dihydrothiachromine species from the **AP** form [43-45]. Nucleophilic attack of N4′ on C2 results in the formation of a C2–N4′ bond, giving rise to the tricyclic intermediate $\mathbf{TCH}^+$. Loss of the N1′ proton $\mathbf{TCH}^+$ will result in the **TC** form of the cofactor [45]. While admittedly not common, the tricyclic form of the cofactor has been observed on at least two ThDP-dependent enzymes. Dihydrothiachromine diphosphate (**TC**) was observed in the X-ray structure of phosphoketolase from *Bifidobacterium breve* [46], and its hydroxyethyl derivative was identified in the structure of acetolactate synthase from *Klebsiella pneumoniae* whose crystals had been soaked with pyruvate [47].

In a very recent study, we used quantum chemical methodology to investigate the detailed reaction mechanism of benzoylformate decarboxylase (BFDC) [29]. A model of the active site was designed on the basis of the X-ray structure of BFDC in complex with the substrate analog inhibitor, (*R*)-mandelate. In that study all intermediates and transition states were located and characterized. Intriguingly, we identified the tricyclic $\mathbf{TCH}^+$ state of the cofactor as an off-cycle intermediate species. It was found to be about 5 kcal/mol lower in energy than the **IP** state, thereby raising the barrier for the formation of the cofactor–substrate adduct (C2α-mandelyl–ThDP). Of course this has important implications for the overall kinetics of any BFDC-catalyzed reaction and, potentially, for all THDP-dependent enzymes [29].

This unexpected result prompted us to conduct a systematic study of the energetics of the various enzyme-bound states of ThDP (Scheme 1). To this end, we have used BFDC as a representative ThDP-dependent enzyme, and employed the quantum chemical approach used to study the BFDC reaction mechanism to characterize the various states of the ThDP cofactor. Models representing different enzymatic and non-enzymatic environments have been generated and, for each model, the cofactor has been characterized in terms of energies and geometries.

Results

The various states of the cofactor have been studied using five different models. In all cases the diphosphate group is omitted since it is thought to act primarily as an anchor for the cofactor and, consequently, was not deemed relevant to the current study. Model **A** is the simplest, representing the cofactor alone in solution. It comprises 31 atoms and has a net charge of +1. Models **B**–**E** represent the cofactor in the BFDC active site in the absence and presence of bound ligands. The active site model is built on the basis of the crystal structure (PDB ID 1MCZ) and is identical to that used in the mechanistic study [29]. As shown in Figure 1, the model comprises all groups that make up the active site pocket, including residues that surround the ThDP cofactor and the ligand. A detailed description of the residues included in the model and the choice of protonation states is provided in reference [29].

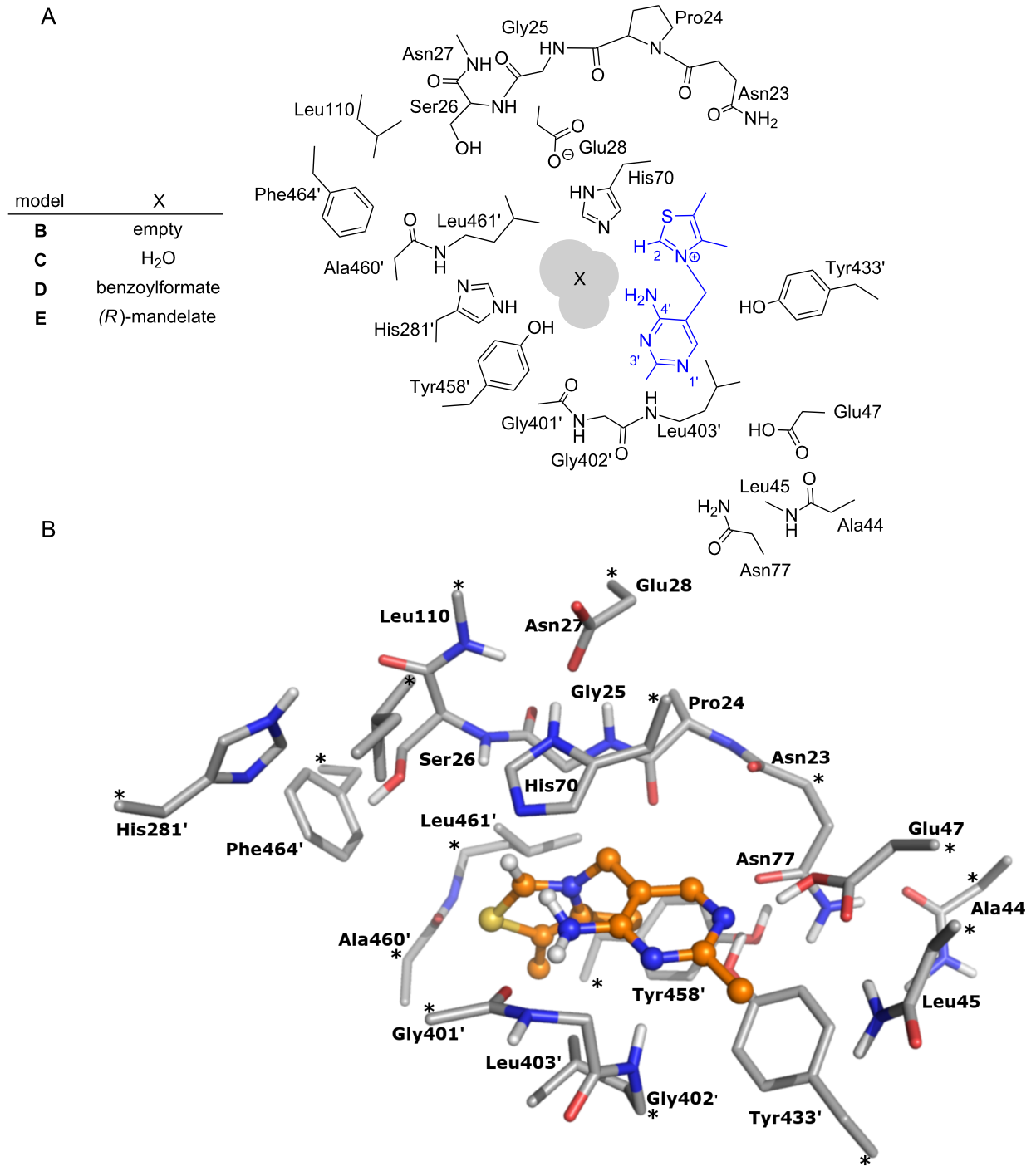


Figure 1: A) 2D representation of ThDP (blue) and the residues included in the active site models, and B) optimized structure of model **B** with an empty active site. Asterisks mark atoms that were kept fixed to their crystallographic positions during the geometry optimization. The BFDC active site has contributions from two monomers and primes re-used to indicate residues from the second monomer. For clarity, the non-polar hydrogens of the residues are not included.

In model **B** the active site does not contain any ligand, and is considered for comparative purposes. It has a total of 291 atoms and a net charge of 0. In model **C** the active site contains a crystallographic water molecule and includes 294 atoms with net charge of 0. In model **D** the water is replaced by the benzoylformate substrate in its deprotonated form and has thus 307 atoms

and a net charge of −1. Finally, in model **E** the active site contains (*R*)-mandelate, again in its deprotonated form, and consists of 309 atoms with a net charge of −1. In the active site models **B**–**E** a number of atoms are kept fixed in the geometry optimizations in order to preserve the overall structure of the active site and avoid excessive movements of the various groups. The fixed atoms are indicated by asterisks in Figure 1B.

The states of the ThDP cofactor considered here are shown in Scheme 1. The starting point for each model is the **AP** state, the energy of which is set to zero, and the energies of the other states are then compared to it.

Model **A**: cofactor in solution

In order to analyze the effect of the enzyme environment on the properties of the various ThDP states, it is important to first consider the solution states of the cofactor in the absence of enzyme. The calculations show that the difference in energy between the lowest energy conformer and the typical V-conformation of enzyme-bound ThDP [48] is 4.2 kcal/mol. Interestingly, the lowest energy structure also adopts a V-shape, but one in which thiazolium ring is perpendicular to the pyrimidine ring (see Supporting Information File 1 for an optimized structure). Given that this study compares enzyme-bound states of ThDP, it is appropriate to use the typical V-conformation of the AP form as the starting/reference point. With that in mind, the optimized geometries of the various V-states of the cofactor alone are displayed in Figure 2.

Calculations on model **A** show that the **AP** state is the most stable, but the tricyclic form **TCH**$^+$ is only 2.8 kcal/mol higher in energy (Table 1). Presumably the proximity of N4' to C2 in the V-conformation makes this state more accessible than it would be if ThDP was unconstrained in solution. Both the **IP** and **YIH**$^+$ states are considerably higher in energy, at +6.4 and +15.2 kcal/mol, respectively. It should be noted that the acid/base conjugates of these states (**APH**$^+$, **TC** and **YI**, respectively) were not calculated, as these structures would have different

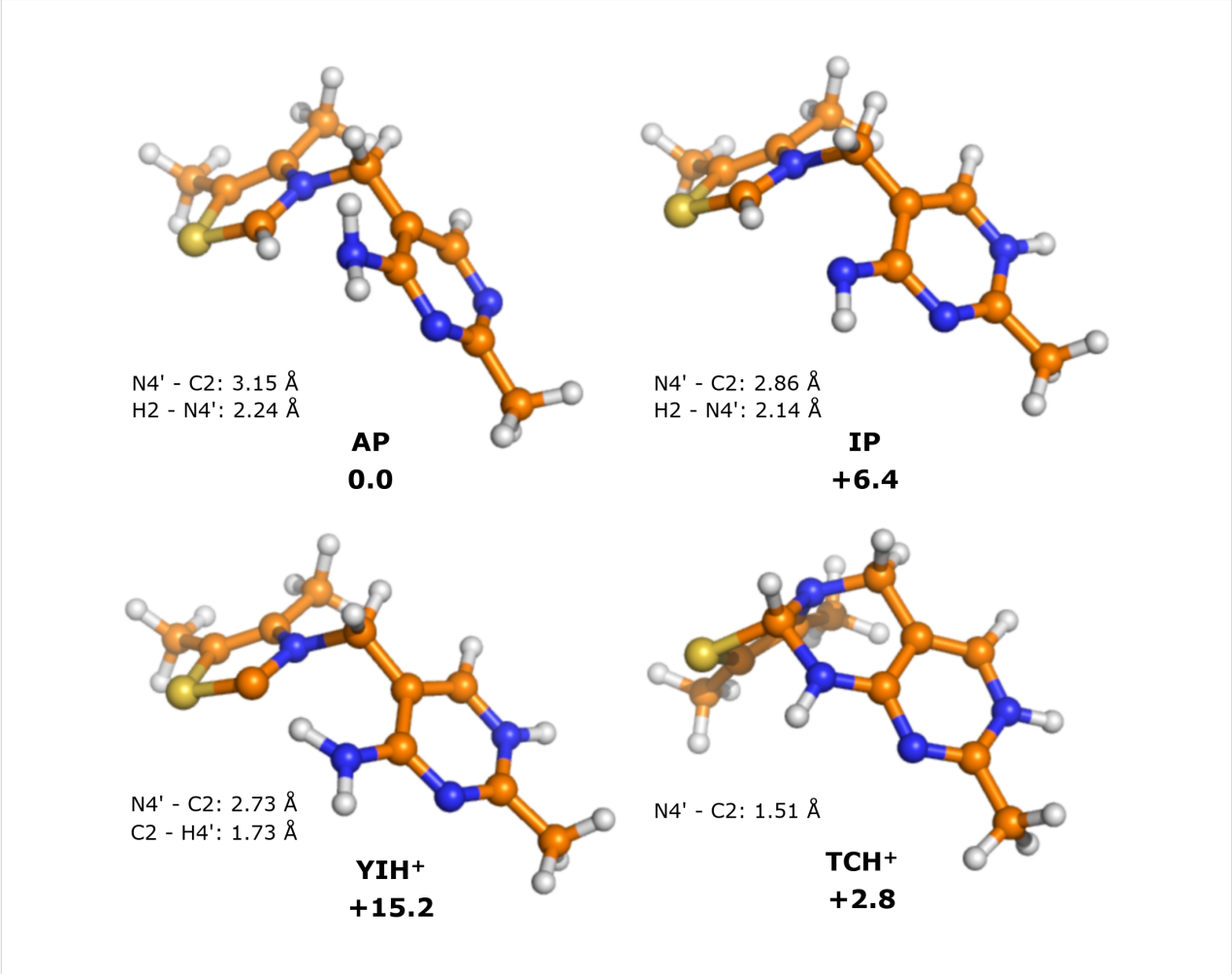


Figure 2: Optimized structures of the states of ThDP in the absence of enzyme (model **A**). Relative energies are indicated in kcal/mol.

numbers of atoms and so the energies would not be directly comparable.

Table 1: Calculated relative energies (kcal/mol) of the various ThDP states. The most stable state for each model is indicated in bold face.

state	model **A**	model **B**	model **C**	model **D**	model **E**
AP	**0.0**	**0.0**	**0.0**	0.0	0.0
APH$^+$	–	+1.2[a]	+1.9[a]	+3.0	**−0.3**
IP	+6.4	+12.8	+16.0	−0.9	+10.2
YI	–	+11.6	+11.0	+3.0	+23.2
YIH$^+$	+15.2	+11.3[a]	+11.1[a]	+6.0	+20.2
TC	–	+2.6	+8.9	−4.9	+7.9
TCH$^+$	+2.8	+2.0[a]	+8.5[a]	**−6.3**	+11.3

[a]Values are calculated with the N1′–H distance constrained to 1.15 Å (see text).

Model **B**: ThDP in the empty active site

Model **B** represents ThDP in the active site of BFDC in the absence of ligand. The geometries of the different ThDP states were optimized (Figure 3) and their energies evaluated (Table 1). Here, the enzyme provides both electrostatic and steric interactions with ThDP, all of which are expected to affect the cofactor's geometry and energy. Of particular interest is the conserved Glu47 residue which forms a hydrogen bond to N1' of the pyrimidine ring. It is important to note that, during the geometry optimizations of the three states **YIH**$^+$, **APH**$^+$ and **TCH**$^+$, the N1' proton invariably transferred spontaneously to the carboxylate of Glu47 thereby yielding the conjugated states **YI**, **AP** and **TC**. In order to assess independently the effect of the N1′ protonation state, approximate energies of **YIH**$^+$, **APH**$^+$ and **TCH**$^+$ were calculated by restraining the N1′–H distance to 1.15 Å. Even with that constraint the energies obtained are within 2 kcal/mol of those of their conjugates (Table 1), showing that N1′ protonation/deprotonation has

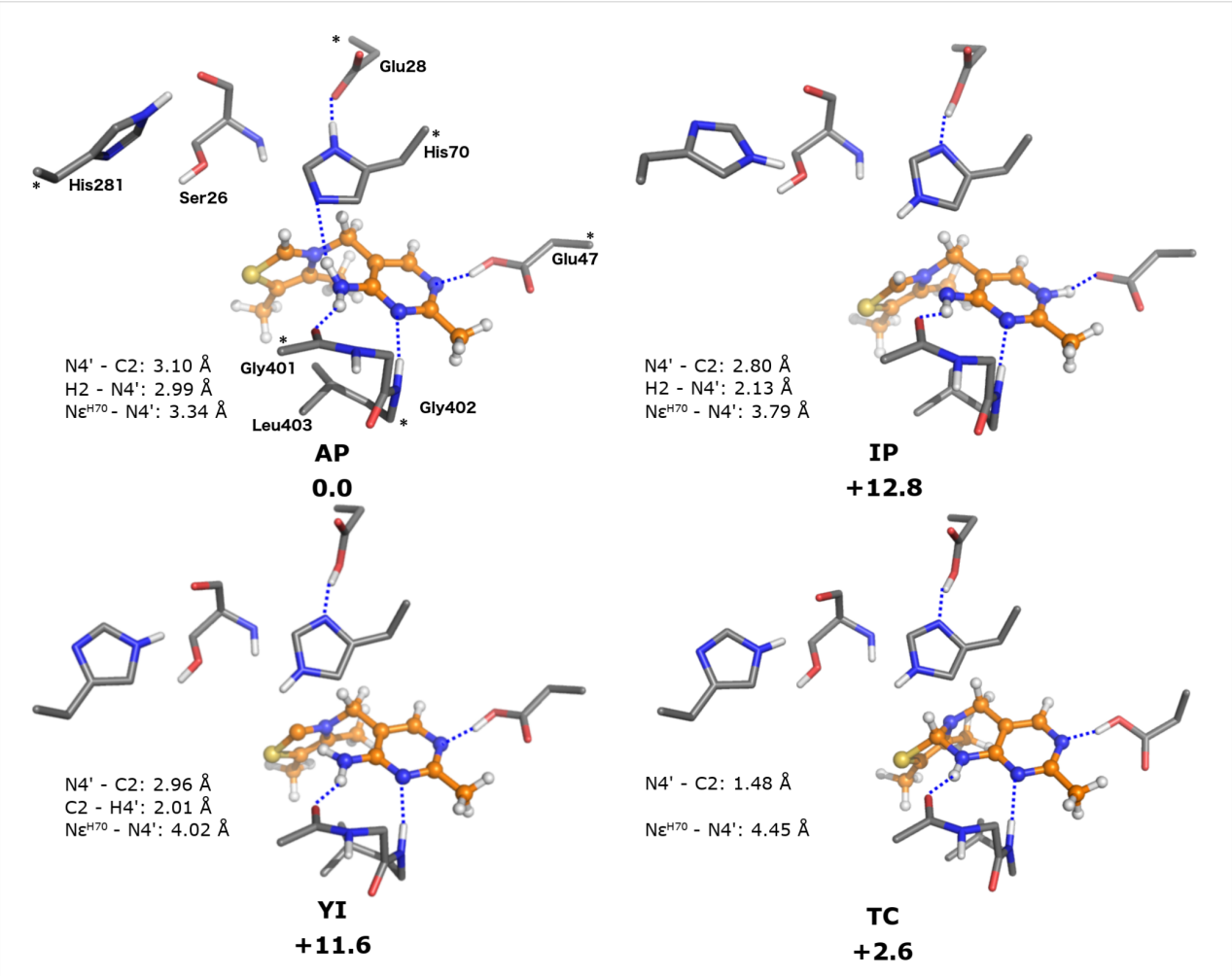


Figure 3: Optimized structures of the states of BFDC-bound ThDP in the absence of ligand (model **B**). Relative energies are indicated in kcal/mol. For clarity, only a selected part of the model is shown, for full model, see Figure 1.

only marginal impact on the relative energies of the cofactor states.

In the **AP** state, His70 interacts with the cofactor through a hydrogen bond between the Nε and the exocyclic N4' amino group, a bond that is not present in the other states. With that notable exception, the overall geometries of the different states are quite similar, and bond distances are also fairly consistent (Figure 3).

As with model **A**, the **AP** state is found to be the lowest energy state in model **B**. The stability of **TC** state in model **B** is similar to that of the $\mathbf{TCH}^+$ in the model of the cofactor alone (+2.6 compared to +2.8 kcal/mol, respectively).The energy of the **YI** state is also reasonably close to that of $\mathbf{YIH}^+$ in the cofactor alone (+11.6 vs +15.2 kcal/mol, relative to their respective **AP** states). Indeed, it was not until the energy of the **IP** state was calculated that the enzyme showed any significant effect. In this instance the **IP** state was calculated to be 12.8 kcal/mol higher than **AP**, i.e., more than 6 kcal/mol higher than the value calculated in the absence of enzyme.

Model **C**: active site including the crystallographic water

In the X-ray structure of unliganded BFDC, there is a crystallographic water molecule that is displaced when a ligand is present [13,49]. In model **C**, that water molecule is included and is found to bind in the same position regardless of the state of the cofactor. A superposition with the crystal structure without substrate (PDB 1BFD) shows that the model calculations reproduce very well the position of this water (see Supporting Information File 1), even though the hydrogen-bonding pattern of the water molecule changes somewhat between the states (Figure 4). Interestingly, there is a hydrogen bond from the

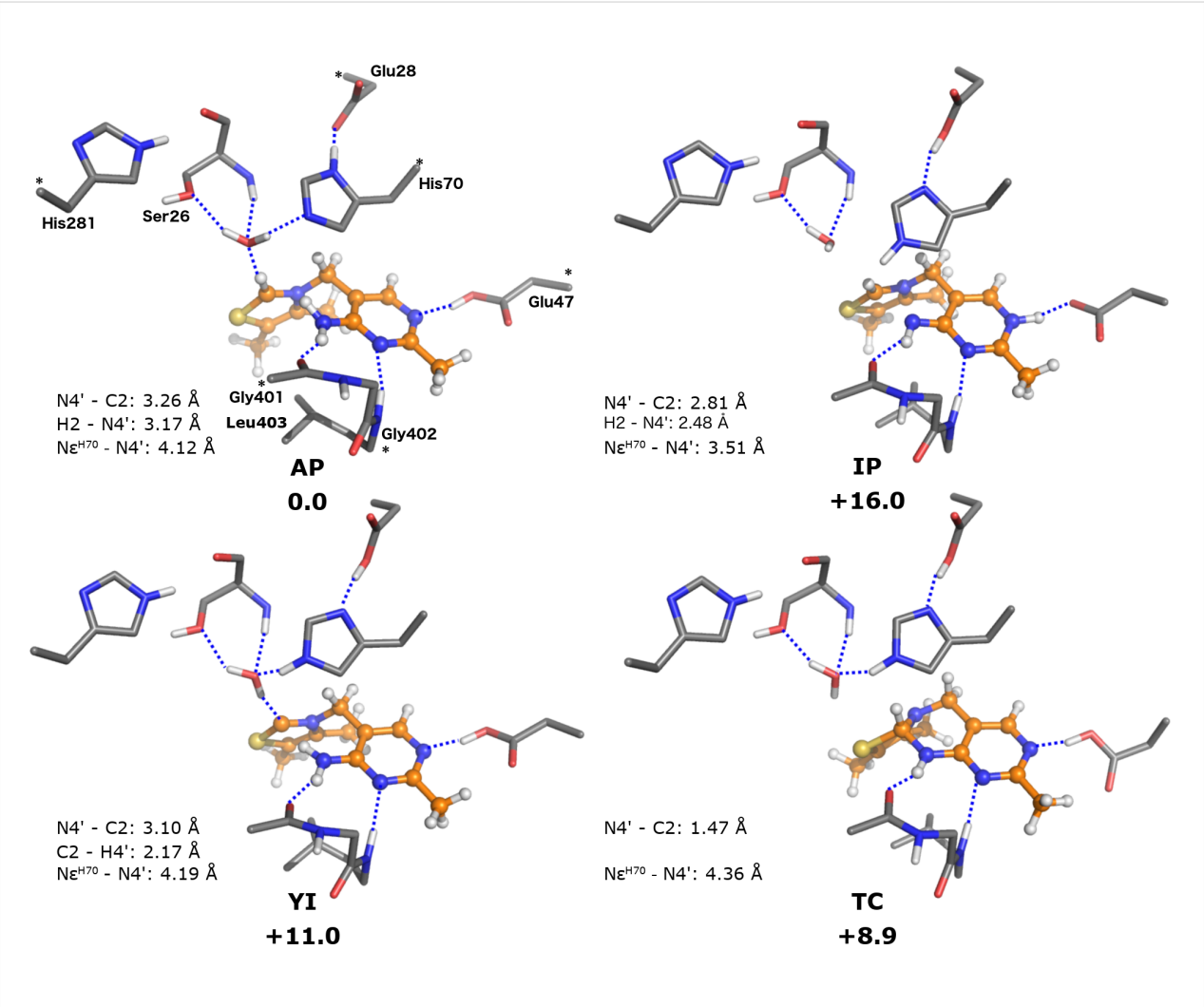


Figure 4: Optimized structures of the ThDP states for the model including the crystallographic water (model **C**). Relative energies are indicated in kcal/mol.

water to the anionic C2 carbon in the **YI** state (OH···C distance of 1.91 Å), with the negative charge on C2 further stabilized by interaction with the exocyclic NH_2 group (NH···C distance of 2.17 Å). The superposition of the structures of model **B** and model **C** reveals that inclusion of the water molecule causes also a slight movement of the thiazolium ring of the cofactor towards the interior of the active site cavity (see Supporting Information File 1).

Energetically, we note that the **IP** state is destabilized compared to that in model **B**, now being 16.0 kcal/mol higher than **AP**, an increase of 3.2 kcal/mol. Conversely, the stability of the **YI** state was very similar (+11.0 vs +11.6 kcal/mol) to that observed for model **B** suggesting the water molecule has little effect on the stability of the ylide. However, the water molecule reserves its largest effect for the **TC** state which now is 8.9 kcal/mol less stable than **AP**, an increase of 6.3 kcal/mol over that observed in model **B**. Overall, it would appear that the effect of the water molecule is to stabilize the **AP** state compared to the other states. The exception is the **YI** state, which seemingly benefits from the new hydrogen bond from the water molecule to the C2 carbanion.

Finally, as with model **B**, the geometries for the $\mathbf{YIH^+}$, $\mathbf{APH^+}$ and $\mathbf{TCH^+}$ states could not be obtained, as geometry optimizations lead to their respective conjugates. However, constrained optimizations again show that the energies are not affected significantly by the protonation (see Table 1).

Model **D**: active site including benzoylformate

In model **D**, which includes the native substrate, benzoylformate (BF), all the states shown in Scheme 1 could be located by the geometry optimizations. As with the crystallographic water in model **C**, the presence of the substrate pushes the thiazolium ring somewhat towards the interior of the cavity. In all ThDP states the carboxylate of BF forms hydrogen bonds to the side chain hydroxy group and the backbone NH of Ser26, and to the Nε of His281 (Figure 5). In the **AP** and $\mathbf{APH^+}$ states, the Nδ of His70 accepts a hydrogen bond from the exocyclic NH_2 group, with an N···HN distance of 2.1 Å. In the other states, the Nε of His70 is protonated and donates a hydrogen bond to the carbonyl of the substrate. In **YI** and $\mathbf{YIH^+}$, the exocyclic NH_2 interacts with the C2 carbanion of the thiazolium ring.

Strikingly, the presence of the substrate has a dramatic impact on the relative stabilities of the various states as compared to the water (model **C**) or the empty cavity (model **B**). Presumably this effect is primarily due to the overall negative charge of the benzoylformate and the bulk of the phenyl substituent. The most significant changes are seen in the energies of the two catalytically productive states, **IP** and **YI**. Now, the former is more stable, by 0.9 kcal/mol, than the **AP** state. This may not seem much but the overall change is substantial as the **IP** state was calculated to be 6.4, 12.8 and 16.0 kcal/mol higher in energy than the **AP** state in models **A**, **B**, and **C**, respectively. The energy of the ylide is also lowered in the presence of the substrate and the **YI** state is now only +3.0 kcal/mol compared to **AP**. In the other models the difference was more than 11 kcal/mol (Table 1).

Although these results clearly suggest that substrate binding results in catalytically productive states of the cofactor, this is not the whole story. Model **D** also indicates that substrate binding produces a major stabilization of the two non-productive tricyclic species. In fact, the most stable state is found to be $\mathbf{TCH^+}$, which is calculated to be 6.3 kcal/mol more stable than the **AP** state. Also the deprotonated **TC** state is 4.9 kcal/mol more stable than **AP**, and both tricyclic states are at least 4 kcal/mol lower in energy than the **IP** form. While substrate binding favoring the non-productive species is surprising and seems counterintuitive, benzoylformate binding also makes the catalytically essential **IP** and **YI** forms more accessible than in any of the other models. Importantly, as detailed in our recent paper on the reaction mechanism of BFDC, this model is consistent with the kinetics of the BFDC reaction [29].

In a final note on model **D**, although the calculations show that proton transfer from N1' to Glu47 is not spontaneous in this model, the energy difference between the conjugated pairs $\mathbf{AP/APH^+}$, $\mathbf{TC/TCH^+}$, $\mathbf{YI/YIH^+}$ remains very low, suggesting the forms are readily interchangeable (Table 1).

Model **E**: active site of BFDC with (*R*)-mandelate bound

In model **E**, in which the active site of BFDC contains the inhibitor (*R*)-mandelate, the hydrogen-bonding network is very similar to that of benzoylformate in model **D**. However, as shown in Figure 6, the benzylic hydroxy group provides a source of additional interactions. In the **AP** and $\mathbf{APH^+}$ states, the hydroxy group forms hydrogen bonds with His70 and the exocyclic NH_2. Support was lent to the validity of model **E** when superposition of the structure of the $\mathbf{APH^+}$ form on the structure of BFDC:(*R*)-mandelate complex (PDB 1MCZ) showed no major movements (see Supporting Information File 1). In the other states, the bond to His 70 is maintained but that to the exocyclic NH2 is broken. Instead the hydroxy group forms a hydrogen bond to the backbone carbonyl of Gly401.

Energetically, we note that, in this model, $\mathbf{AP/APH^+}$ are by far the most stable states with the **IP**, **YI** and **TC** states being 10.2, 23.2 and 7.9 kcal/mol higher than **AP**, respectively (Table 1).

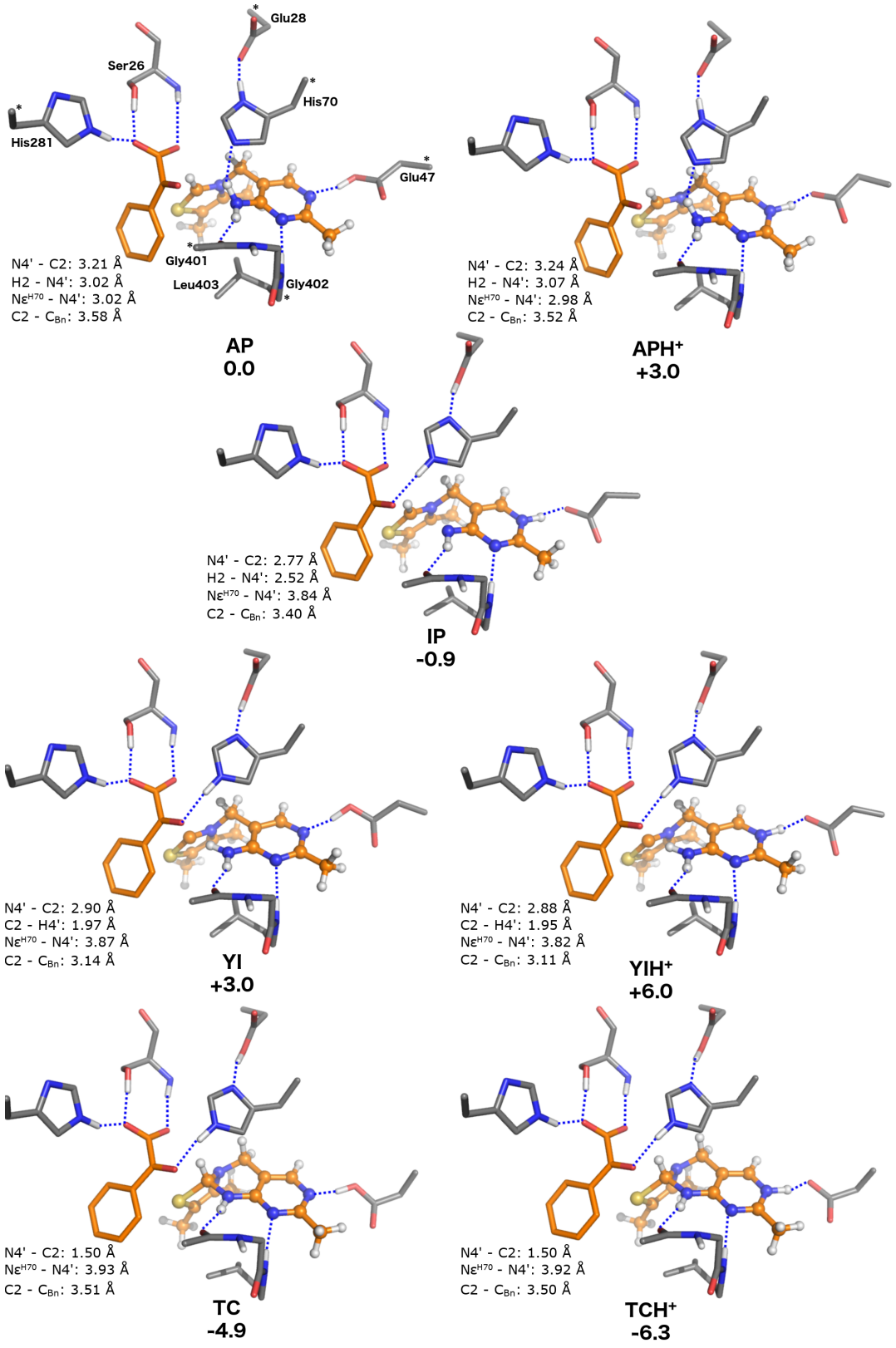


Figure 5: Optimized structures of the ThDP states in the BFDC active site containing the substrate, benzoylformate (model **D**). Relative energies are indicated in kcal/mol.

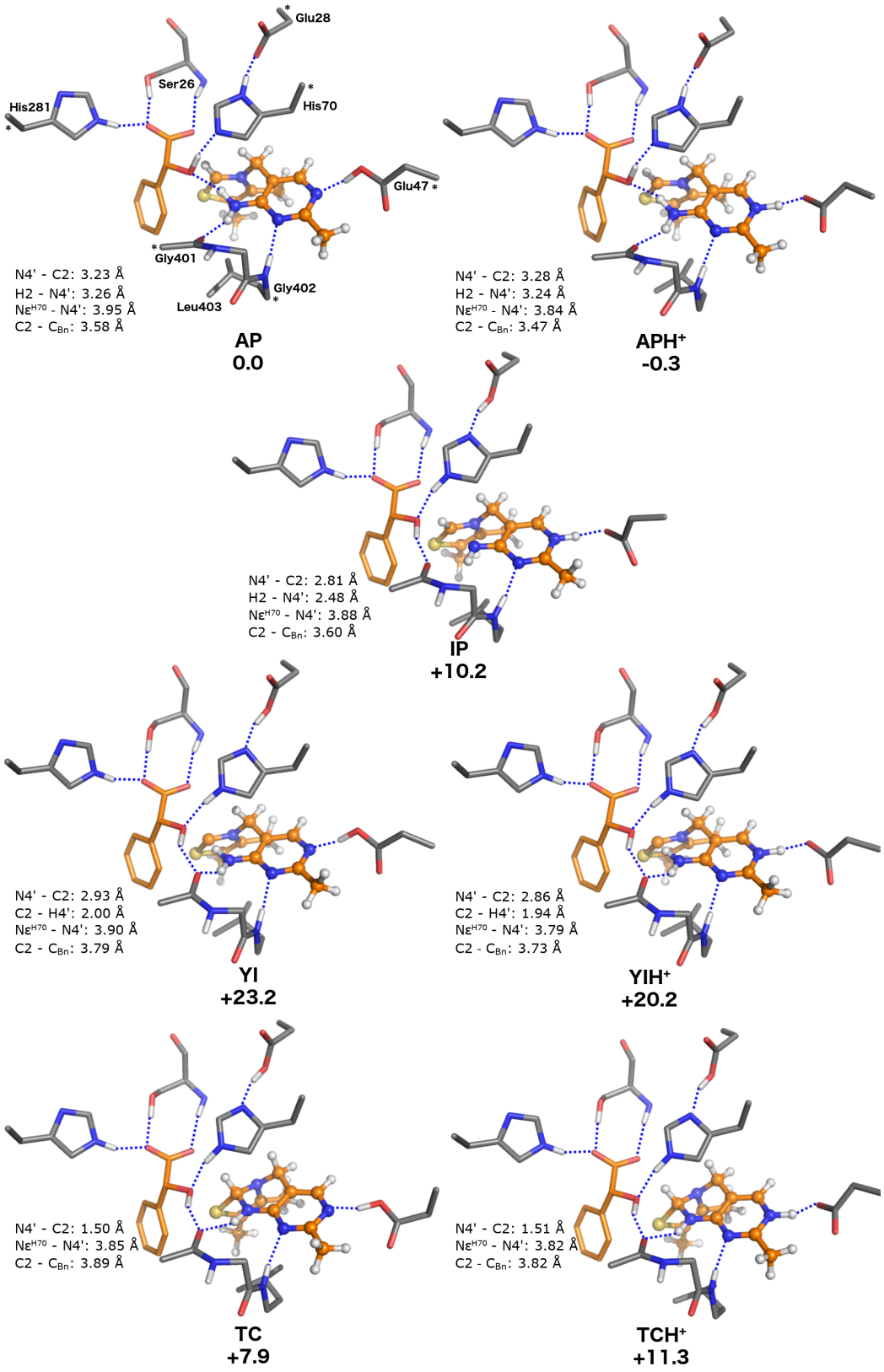


Figure 6: Optimized structures of the ThDP states for the model including (*R*)-mandelate (model **E**). Relative energies are indicated in kcal/mol.

Further, each of these states is more than 10 kcal/mol higher in energy relative to the **AP** state than its counterpart in model **D** in which the substrate is bound (Table 1). Clearly the binding of the (*R*)-mandelate causes a stabilization of the **AP**/**APH**$^+$ states relative to the others.

As seen from Table 1, despite having the same overall charge and similar bulkiness of the substituents, the binding of the benzoylformate (model **D**) and (*R*)-mandelate (model **E**) result in quite different energies. The superposition of the **AP** states of the two models (see Supporting Information File 1) shows that the additional hydrogen bond provided by the benzylic hydroxy group of (*R*)-mandelate (vide supra) contributes significantly to this difference. Further, changing from the sp^2 carbonyl carbon to the sp^3 benzylic carbon results in a substantial movement of the substituent oxygen which also contributes to the energy difference between the models.

According to Table 1 the **AP**/**APH**$^+$ forms are the most stable states for models **C** and **E**, i.e., BFDC in the absence of ligand and in the presence of (*R*)-mandelate. The CD spectrum of BFDC shows a small minimum at around 325 nm, attributed to the **AP** form. Titrating BFDC with methyl benzoylphosphonate (MBP), a mechanism-based inhibitor, gave rise to a new maximum at around 300 nm, attributed to the **IP** form, with a concurrent loss of signal at 325 nm [50]. Based on the data in Table 1, it is not certain that the titration with benzoylformate would give rise to the **IP** form. However, the data unambiguously suggest that titration of BFDC with (*R*)-mandelate should result in no change in its CD spectrum, even when the enzyme is saturated. Accordingly, the titration was carried out and, indeed, even at (*R*)-mandelate concentrations well in excess of its K_i value of 1 mM [49], no change in the spectrum was observed (see Supporting Information File 1 for details).

Discussion

Since it was first purified from beer yeast over 80 years ago [51], the structure of ThDP and its related intermediates and ionization states have undergone intensive investigation (summarized in references [2,52]). Most of these investigations have focused on the structure and properties of the covalently modified ThDP intermediates of diverse ThDP-dependent enzymes. Less attention has been paid to the variety of states the cofactor itself can adopt on the enzyme. As shown in Scheme 1, when various tautomers and ionization states are included, ThDP can adopt at least seven forms on any given ThDP-dependent enzyme. This is prior to any reaction taking place. Most of these are accounted for in typical analyses [52] but the two tricyclic forms, **TC** and **TCH**$^+$, are rarely considered. Unlike the tricyclic states, which could be regarded to be non-productive, the **IP** and the **YI** forms arising from it are essential for catalysis and always considered in any mechanistic study. However, conceivably all seven states could be energetically accessible and could influence the catalytic mechanism. Over the past several years there has been an increasing use of ThDP-dependent enzymes as chiral catalysts [53]. Given that all of these enzymes will require ready access to the **IP** and **YI** forms, it seemed logical to take a closer look at the relative energies of the various states and how those energies may be affected by the binding of different ligands. Toward that end, we have used DFT calculations to explore the energetics of the various states of enzyme-bound ThDP using benzoylformate decarboxylase as the model enzyme.

The work was predicated on two elements. First, that the cofactor was held in a V-conformation on the enzyme, and second, that the resting (reference) state of the cofactor was the **AP** form. Both have been confirmed experimentally [13,49,50] and are typical of most, if not all, of the ThDP-dependent enzymes studied to date. In total, five models were employed: models **A** and **B** providing a comparison of the cofactor states in the presence and absence of enzyme, and models **C**–**E** examining the effects of active site ligands.

Models **A** and **B** both predict the **AP** state to be the most stable, vindicating its use as the reference state. Perhaps the first surprise was the difference in magnitude and overall effect the active site ligands had on the relative energy levels. For example, the simple addition of a crystallographic water destabilized both the **IP** and **TC** forms by 3 and 6 kcal/mol, respectively, thereby ensuring that BFDC largely exists as the **AP** form. Even more surprising was the comparison of the substrate, benzoylformate, and the inhibitor, (*R*)-mandelate. With the exception of an sp^3 rather than sp^2 hybridized benzylic carbon, (*R*)-mandelate is identical to benzoylformate. However, they have markedly different effects on the states of the cofactor. In model **C**, corresponding to the native enzyme, the **IP** form is 16 kcal/mol less stable than the **AP** form. When benzoylformate binds (model **D**), the **IP** form becomes energetically favored by 0.9 kcal/mol, an overall change of 17 kcal/mol. This is accompanied by an 8 kcal/mol stabilization of the catalytically essential **YI** form. Conversely, when the inhibitor is bound (model **E**), the **IP** and **YI** forms are ca. 10 and 20 kcal/mol less stable than in model **D**. Clearly the substrate-induced changes combine to facilitate catalysis, while those brought about by the inhibitor make reaction more difficult.

Considering that they have been largely ignored in previous studies, the next surprise was that the tricyclic forms **TC** and **TCH**$^+$ were relatively stable, in both the absence and presence of enzyme. In fact, it seemed that the primary effect of the binding of ThDP to the enzyme was to bring about the destabi-

lization of the **IP** form. Over all five models, the tricyclic forms were consistently more energetically stable than the catalytically essential **IP** and **YI** forms (Table 1). While it may be argued that stabilization of the tricyclic forms could prove to be enzyme specific, the relatively low energy of the tricyclic state in the absence of enzyme cannot be disregarded, and certainly suggests that the **TC**/**TCH**$^+$ forms may be more common than previously recognized. Further, and consistent with results for the **YI**/**YIH**$^+$ forms, the relative stabilities of **TC**/**TCH**$^+$ states proved to be ligand specific. In model **C**, the **AP** form is ca. 9 kcal/mol more stable than the **TC**/**TCH**$^+$ forms. However, after substrate binding (model **D**) the **TC**/**TCH**$^+$ states are ca. 5 kcal/mol more stable than **AP**. Thus, the presence of benzoylformate shifts the relative energies by 14 kcal/mol and, concomitantly, makes the tricyclic forms the most stable species. The inhibitor again provides the contrast, for the binding of (*R*)-mandelate (model **E**) has virtually no effect on the relative energy levels of the tricyclic forms, and **AP** remains clearly the most stable state.

At this point, it is reasonable to assess the validity of the current computational results in light of available experimental information. In the first instance the results confirm that the **AP** state is the lowest energy, i.e., resting state. This was one of the elements on which the work was predicated and is consistent with data obtained from, among others, BFDC, benzaldehyde lyase, pyruvate oxidase, pyruvate decarboxylase and the E1 subunit of the pyruvate dehydrogenase complex (summarized in [54]). In fact, there are only two cases in which the **IP** form has been observed in the resting enzyme, namely pyruvate oxidase and the pyruvate dehydrogenase complex. In both cases, the **AP** state was the predominant form [54].

Secondly, the results show that substrate binding dramatically lowers the energy of the **IP** and **YI** states, which would, presumably, increase the rate and extent of ylide formation. This observation is more difficult to demonstrate experimentally. While H/D exchange experiments have been used as a measure of the rate of ylide formation, substrate activation has only been observed with allosteric enzymes such as yeast PDC [8]. Further, even though there is a CD signature for the **IP** state, it is usually associated with formation of a tetrahedral reaction intermediate. As a result it is difficult to separate any increase in the **IP** signal arising from substrate binding from that due to intermediate formation. Possibly the closest to experimental support came from an experiment in which the reaction of BFDC with MBP was monitored by stopped-flow measurements at 308 nm using the intrinsic absorbance of the **IP** state. In that case the results implied that there was a transient formation of a Michaelis complex which was accompanied by an increase in the **IP** form [55].

Next, the calculations suggest that the binding of (*R*)-mandelate should not change the state of the cofactor. Again, somewhat difficult to prove conclusively but titration of BFDC with the substrate analogue, MBP, provided clear evidence for the conversion of the **AP** to the **IP** state for the former. Conversely, and consistent with predictions, the **AP** state remained unchanged when a similar titration was carried out with the inhibitor, (*R*)-mandelate (see Supporting Information File 1).

Finally, what evidence is there for the formation of tricyclic states? As noted in the introduction, there has been little or no effort to identify tricyclic intermediates on ThDP-dependent enzymes. Critically, even though they were treated dismissively, there are two X-ray structures which, at a minimum, provide unambiguous evidence for the formation of stable tricyclic intermediates on an enzyme [46,47]. Additional evidence, albeit more indirect, comes from an inhibition study using omeprazole, which was predicted to possibly interact with ThDP-dependent enzymes. The prediction was based on the similarity of omeprazole to the tricyclic form of thiamin. This was confirmed experimentally when omeprazole was subsequently shown to be a competitive inhibitor of both transketolase and PDC, with a K_i value for the latter only ca. 20 times the K_m for ThDP measured in the same experiment [56].

The current calculations show that, in the presence of substrate, **TC**/**TCH**$^+$ are the most stable states of ThDP on BFDC. Yet, even though a large number of high-resolution structures of BFDC variants, in the presence and absence of ligands, have been determined, none of them shows the tricyclic intermediate. This may seem surprising but it must be considered that when benzoylformate is present the catalytic cycle is in operation, reactions are running and covalent ThDP intermediates are being formed. As detailed in our recent paper, the enamine is the most stable reaction intermediate [29], which makes it unlikely that the **TC** state will be detected experimentally. Furthermore, in the absence of the substrate (model **C**) or in the presence of the inhibitor (model **E**), the **TC** states are clearly disfavored, with calculated energies of +8.9 and +7.9 kcal/mol, respectively, relative to the **AP** state.

Are the tricyclic forms even relevant? That is really the crux of the matter, and the answer is, for BFDC at least, yes! It is important to note that the tricyclic forms **TC**/**TCH**$^+$ are calculated to be more stable than the ylide forms **YI**/**YIH**$^+$ in all considered models. Further, the calculations on model **D** indicate that an energy penalty of ca. 5 kcal/mol must be paid to go from the **TCH**$^+$ to the **YI** state, which is the catalytically active form of the cofactor. This, in turn, effectively increases the barrier for formation of the first reaction intermediate. In fact, as shown in the paper on the catalytic mechanism that inspired

this work, the energy barrier brought about by the stable tricyclic state fits well with the experimental evidence for the slow first step, i.e., formation of the mandelylThDP adduct [29].

In addition to BFDC, the X-ray structures of tricyclic intermediates suggest that an even greater stabilization is present on phosphoketolase and acetolactate synthase. It could well be argued that **TC** stabilization may prove to be rare and specific to only a few enzymes. Yet, the relatively low energy of the tricyclic state in the absence of enzyme cannot be disregarded, and certainly suggests that the **TC** forms may be more common than previously thought. Over the past few years, a rapid-quench NMR technique has been employed to determine microscopic rate constants for elementary steps in several ThDP-dependent enzymes. It is notable that, in addition to BFDC [55], *E. coli* AHAS I and II [57], glyoxylate carboligase [7], DXP synthase [58] and indolepyruvate decarboxylase [59] all have formation of the first tetrahedral intermediate as the rate-determining step. Of course, in the absence of the corresponding calculations it is impossible to definitively state that this is due to stabilization of the **TC** state, but the question is worth asking. The pyruvate oxidase from *Lactobacillus plantarum* provides some support in that both the **AP** and **IP** forms are present in the resting enzyme [54] and decarboxylation, rather than formation of the first intermediate, was found to be rate limiting [60]. On the other hand, product release was the slowest step for *Zm*PDC and *Sc*PDC [61], so clearly not all ThDP-dependent enzymes behave in the same manner.

Of course, while the relative stability of the **TC** form may slow down the BFDC reaction, it is conceivable that it may also play a beneficial role. As the pK_a of the C2 proton decreases, the activity of the ThDP cofactor increases [62]. However, concomitantly, the thiazolium ring becomes more susceptible to hydrolysis to a catalytically inactive form [44]. The stable tricyclic form of the cofactor, which can readily revert to its active form, may provide a protective mechanism against hydrolysis [44].

Two final thoughts: first, the current results show that even when substrate is bound, the tricyclic state, not the ylide, is the most energetically stable. This observation implies that starting the computational investigations of the ThDP-dependent catalytic mechanism directly from the ylide, as done in numerous examples in the literature, may give rise to an incomplete, if not inaccurate, picture of the energy profile of the reaction. Second, many ThDP-dependent enzymes are being evaluated for use as biocatalysts. The stark difference in the effect of two very similar ligands, benzoylformate and (*R*)-mandelate, on the activation of the cofactor suggests that the use of alternative substrates or, possibly more importantly, the evolution of ThDP-dependent enzymes to accept a wide range of non-native substrates, might not be as simple as may have been expected.

Experimental

All calculations were performed with the B3LYP-D3(BJ) [62-65] density functional method and using the Gaussian 09 package [66]. The geometries were optimized with the 6-31G(d,p) basis set, and the energy of the stationary points was refined by single-point calculations with 6-311+G(2d,2p) basis set. Frequency calculations were done at the same level of theory as the optimizations to obtain zero-point energy corrections, and solvation energies were calculated using the implicit solvent method SMD [67] with a dielectric constant ε = 4.

Supporting Information

Supporting Information File 1

Lowest-energy conformation of model **A**, superpositions of the **AP** state/model **C** and **APH**$^+$ state/model **E** with crystal structures, superposition of the **AP** state/models **B** and **C**, superposition of the **AP** states/models **D** and **E**, experimental CD spectra, calculated energies and energy corrections, and Cartesian coordinates of all optimized structures.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-15-S1.pdf]

Acknowledgements

FH acknowledges financial support from the Swedish Research Council and MJM appreciates the support of the National Science Foundation (CHE 1306877).

ORCID® iDs

Ferran Planas - https://orcid.org/0000-0002-5070-9988
Michael J. McLeish - https://orcid.org/0000-0002-9221-9119
Fahmi Himo - https://orcid.org/0000-0002-1012-5611

References

1. Pohl, M.; Sprenger, G. A.; Müller, M. *Curr. Opin. Biotechnol.* **2004,** *15,* 335–342. doi:10.1016/j.copbio.2004.06.002
2. Kluger, R.; Tittmann, K. *Chem. Rev.* **2008,** *108,* 1797–1833. doi:10.1021/cr068444m
3. Müller, M.; Sprenger, G. A.; Pohl, M. *Curr. Opin. Chem. Biol.* **2013,** *17,* 261–270. doi:10.1016/j.cbpa.2013.02.017
4. Alvarez, F. J.; Ermer, J.; Huebner, G.; Schellenberger, A.; Schowen, R. L. *J. Am. Chem. Soc.* **1991,** *113,* 8402–8409. doi:10.1021/ja00022a030
5. Breslow, R. *J. Am. Chem. Soc.* **1958,** *80,* 3719–3726. doi:10.1021/ja01547a064
6. Nemeria, N.; Korotchkina, L.; McLeish, M. J.; Kenyon, G. L.; Patel, M. S.; Jordan, F. *Biochemistry* **2007,** *46,* 10739–10744. doi:10.1021/bi700838q

7. Kaplun, A.; Binshtein, E.; Vyazmensky, M.; Steinmetz, A.; Barak, Z.; Chipman, D. M.; Tittmann, K.; Shaanan, B. *Nat. Chem. Biol.* **2008,** *4,* 113–118. doi:10.1038/nchembio.62
8. Kern, D.; Kern, G.; Neef, H.; Tittmann, K.; Killenberg-Jabs, M.; Wikner, C.; Schneider, G.; Hübner, G. *Science* **1997,** *275,* 67–70. doi:10.1126/science.275.5296.67
9. Furey, W.; Arjunan, P.; Chen, L.; Sax, M.; Guo, F.; Jordan, F. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1998,** *1385,* 253–270. doi:10.1016/s0167-4838(98)00073-9
10. Lindqvist, Y.; Schneider, G.; Ermler, U.; Sundström, M. *EMBO J.* **1992,** *11,* 2373–2379. doi:10.1002/j.1460-2075.1992.tb05301.x
11. Muller, Y. A.; Lindqvist, Y.; Furey, W.; Schulz, G. E.; Jordan, F.; Schneider, G. *Structure* **1993,** *1,* 95–103. doi:10.1016/0969-2126(93)90025-c
12. Arjunan, P.; Umland, T.; Dyda, F.; Swaminathan, S.; Furey, W.; Sax, M.; Farrenkopf, B.; Gao, Y.; Zhang, D.; Jordan, F. *J. Mol. Biol.* **1996,** *256,* 590–600. doi:10.1006/jmbi.1996.0111
13. Hasson, M. S.; Muscate, A.; McLeish, M. J.; Polovnikova, L. S.; Gerlt, J. A.; Kenyon, G. L.; Petsko, G. A.; Ringe, D. *Biochemistry* **1998,** *37,* 9918–9930. doi:10.1021/bi973047e
14. Mosbacher, T. G.; Mueller, M.; Schulz, G. E. *FEBS J.* **2005,** *272,* 6067–6076. doi:10.1111/j.1742-4658.2005.04998.x
15. Jordan, F.; Zhang, Z.; Sergienko, E. *Bioorg. Chem.* **2002,** *30,* 188–198. doi:10.1006/bioo.2002.1249
16. Kochetov, G. A.; Usmanov, R. A.; Merzlov, V. P. *FEBS Lett.* **1970,** *9,* 265–266. doi:10.1016/0014-5793(70)80372-6
17. Patel, H.; Nemeria, N. S.; Andrews, F. H.; McLeish, M. J.; Jordan, F. *Biochemistry* **2014,** *53,* 2145–2152. doi:10.1021/bi4015743
18. Balakrishnan, A.; Paramasivam, S.; Chakraborty, S.; Polenova, T.; Jordan, F. *J. Am. Chem. Soc.* **2012,** *134,* 665–672. doi:10.1021/ja209856x
19. Jordan, F. *J. Am. Chem. Soc.* **1974,** *96,* 3623–3630. doi:10.1021/ja00818a041
20. Jordan, F. *J. Am. Chem. Soc.* **1976,** *98,* 808–813. doi:10.1021/ja00419a030
21. Friedemann, R.; Tittmann, K.; Golbik, R.; Hübner, G. *Int. J. Quantum Chem.* **2004,** *99,* 109–114. doi:10.1002/qua.20132
22. DuPré, D. B.; Wong, J. L. *J. Phys. Chem. A* **2007,** *111,* 2172–2181. doi:10.1021/jp067308i
23. Delgado, E. J.; Alderete, J. B.; Jaña, G. A. *J. Mol. Model.* **2011,** *17,* 2735–2739. doi:10.1007/s00894-011-1076-6
24. Jaña, G. A.; Delgado, E. J. *J. Mol. Model.* **2013,** *19,* 3799–3803. doi:10.1007/s00894-013-1908-7
25. Lie, M. A.; Celik, L.; Jørgensen, K. A.; Schiøtt, B. *Biochemistry* **2005,** *44,* 14792–14806. doi:10.1021/bi051134y
26. Wang, J.; Dong, H.; Li, S.; He, H. *J. Phys. Chem. B* **2005,** *109,* 18664–18672. doi:10.1021/jp052802s
27. Alstrup Lie, M.; Schiøtt, B. *J. Comput. Chem.* **2008,** *29,* 1037–1047. doi:10.1002/jcc.20860
28. Hou, Q.; Gao, J.; Liu, Y.; Liu, C. *Theor. Chem. Acc.* **2012,** *131,* 1280. doi:10.1007/s00214-012-1280-1
29. Planas, F.; Sheng, X.; McLeish, M. J.; Himo, F. *Front. Chem. (Lausanne, Switz.)* **2018,** *6,* No. 205. doi:10.3389/fchem.2018.00205
30. Topal, K. G.; Atilgan, C.; Demir, A. S.; Aviyente, V. *Biopolymers* **2010,** *93,* 32–46. doi:10.1002/bip.21291
31. Sánchez, L.; Jaña, G. A.; Delgado, E. J. *J. Comput. Chem.* **2014,** *35,* 488–494. doi:10.1002/jcc.23523
32. Jaña, G.; Jiménez, V.; Villà-Freixa, J.; Prat-Resina, X.; Delgado, E.; Alderete, J. *Proteins: Struct., Funct., Bioinf.* **2010,** *78,* 1774–1788. doi:10.1002/prot.22693
33. Xiong, Y.; Liu, J.; Yang, G.-F.; Zhan, C.-G. *J. Comput. Chem.* **2010,** *31,* 1592–1602. doi:10.1002/jcc.21356
34. Alvarado, O.; Jaña, G.; Delgado, E. J. *J. Comput.-Aided Mol. Des.* **2012,** *26,* 977–982. doi:10.1007/s10822-012-9589-3
35. Lizana, I.; Jaña, G. A.; Delgado, E. J. *J. Chem. Inf. Model.* **2015,** *55,* 1640–1644. doi:10.1021/acs.jcim.5b00197
36. Sheng, X.; Liu, Y. *Biochemistry* **2013,** *52,* 8079–8093. doi:10.1021/bi400577f
37. Zhang, J.; Sheng, X.; Hou, Q.; Liu, Y. *Int. J. Quantum Chem.* **2014,** *114,* 375–382. doi:10.1002/qua.24573
38. Zhu, W.; Liu, Y. *Theor. Chem. Acc.* **2014,** *133,* 1442. doi:10.1007/s00214-013-1442-9
39. Sheng, X.; Liu, Y.; Zhang, R. *RSC Adv.* **2014,** *4,* 35777–35788. doi:10.1039/c4ra03611e
40. White, J. K.; Handa, S.; Vankayala, S. L.; Merkler, D. J.; Woodcock, H. L. *J. Phys. Chem. B* **2016,** *120,* 9922–9934. doi:10.1021/acs.jpcb.6b07248
41. Nauton, L.; Hélaine, V.; Théry, V.; Hecquet, L. *Biochemistry* **2016,** *55,* 2144–2152. doi:10.1021/acs.biochem.5b00787
42. Sheng, X.; Liu, Y.; Liu, C. *J. Mol. Graphics Mod.* **2013,** *39,* 23–28. doi:10.1016/j.jmgm.2012.11.001
43. Metzler, D. E.; Maier, G. D. *Ann. N. Y. Acad. Sci.* **1962,** *98,* 495–498. doi:10.1111/j.1749-6632.1962.tb30570.x
44. Doughty, M. B.; Lawrence, D. S. *J. Chem. Soc., Chem. Commun.* **1985,** 454–455. doi:10.1039/c39850000454
45. Washabaugh, M. W.; Yang, C. C.; Hollenbach, A. D.; Chen, P. *Bioorg. Chem.* **1993,** *21,* 170–191. doi:10.1006/bioo.1993.1016
46. Suzuki, R.; Katayama, T.; Kim, B.-J.; Wakagi, T.; Shoun, H.; Ashida, H.; Yamamoto, K.; Fushinobu, S. *J. Biol. Chem.* **2010,** *285,* 34279–34287. doi:10.1074/jbc.m110.156281
47. Pang, S. S.; Duggleby, R. G.; Schowen, R. L.; Guddat, L. W. *J. Biol. Chem.* **2004,** *279,* 2242–2253. doi:10.1074/jbc.m304038200
48. Guo, F.; Zhang, D.; Kahyaoglu, A.; Farid, R. S.; Jordan, F. *Biochemistry* **1998,** *37,* 13379–13391. doi:10.1021/bi9807097
49. Polovnikova, E. S.; McLeish, M. J.; Sergienko, E. A.; Burgner, J. T.; Anderson, N. L.; Bera, A. K.; Jordan, F.; Kenyon, G. L.; Hasson, M. S. *Biochemistry* **2003,** *42,* 1820–1830. doi:10.1021/bi026490k
50. Brandt, G. S.; Kneen, M. M.; Chakraborty, S.; Baykal, A. T.; Nemeria, N.; Yep, A.; Ruby, D. I.; Petsko, G. A.; Kenyon, G. L.; McLeish, M. J.; Jordan, F.; Ringe, D. *Biochemistry* **2009,** *48,* 3247–3257. doi:10.1021/bi801950k
51. Lohman, K.; Schuster, P. *Biochem. Z.* **1937,** *294,* 188–193.
52. Jordan, F.; Nemeria, N. S. *Bioorg. Chem.* **2014,** *57,* 251–262. doi:10.1016/j.bioorg.2014.08.002
53. Müller, M.; Gocke, D.; Pohl, M. *FEBS J.* **2009,** *276,* 2894–2904. doi:10.1111/j.1742-4658.2009.07017.x
54. Nemeria, N. S.; Chakraborty, S.; Balakrishnan, A.; Jordan, F. *FEBS J.* **2009,** *276,* 2432–3446. doi:10.1111/j.1742-4658.2009.06964.x
55. Bruning, M.; Berheide, M.; Meyer, D.; Golbik, R.; Bartunik, H.; Liese, A.; Tittmann, K. *Biochemistry* **2009,** *48,* 3258–3268. doi:10.1021/bi801957d
56. Nixon, P. F.; Diefenbach, R. J.; Duggleby, R. G. *Biochem. Pharmacol.* **1992,** *44,* 177–179. doi:10.1016/0006-2952(92)90053-l
57. Belenky, I.; Steinmetz, A.; Vyazmensky, M.; Barak, Z.; Tittmann, K.; Chipman, D. M. *FEBS J.* **2012,** *279,* 1967–1979. doi:10.1111/j.1742-4658.2012.08577.x

58. Patel, H.; Nemeria, N. S.; Brammer, L. A.; Freel Meyers, C. L.; Jordan, F. *J. Am. Chem. Soc.* **2012,** *134,* 18374–18379. doi:10.1021/ja307315u
59. Schütz, A.; Golbik, R.; König, S.; Hübner, G.; Tittmann, K. *Biochemistry* **2005,** *44,* 6164–6179. doi:10.1021/bi0473354
60. Wille, G.; Meyer, D.; Steinmetz, A.; Hinze, E.; Golbik, R.; Tittmann, K. *Nat. Chem. Biol.* **2006,** *2,* 324–328. doi:10.1038/nchembio788
61. Tittmann, K.; Golbik, R.; Uhlemann, K.; Khailova, L.; Schneider, G.; Patel, M.; Jordan, F.; Chipman, D. M.; Duggleby, R. G.; Hübner, G. *Biochemistry* **2003,** *42,* 7885–7891. doi:10.1021/bi034465o
62. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988,** *37,* 785–789. doi:10.1103/physrevb.37.785
63. Becke, A. D. *J. Chem. Phys.* **1993,** *98,* 5648–5652. doi:10.1063/1.464913
64. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010,** *132,* 154104. doi:10.1063/1.3382344
65. Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011,** *32,* 1456–1465. doi:10.1002/jcc.21759
66. *Gaussian 09,* Revision D.01; Gaussian Inc.: Wallingford, CT, 2013.
67. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009,** *113,* 6378–6396. doi:10.1021/jp810292n

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Back to the future: Why we need enzymology to build a synthetic metabolism of the future

Tobias J. Erb[1,2]

Review

Open Access

Address:
[1]Max-Planck-Institute for Terrestrial Microbiology, Department of Biochemistry & Synthetic Metabolism, Karl-von-Frisch-Str. 10, D-35043 Marburg, Germany and [2]LOEWE Center for Synthetic Microbiology (SYNMIKRO), Marburg, Germany

Email:
Tobias J. Erb - toerb@mpi-marburg.mpg.de

Keywords:
enzymes; in vitro biochemistry; metabolic engineering; synthetic biology

Beilstein J. Org. Chem. **2019,** *15,* 551–557.
doi:10.3762/bjoc.15.49

Received: 12 October 2018
Accepted: 29 January 2019
Published: 26 February 2019

This article is part of the thematic issue "Enzymes in chemical transformations".

Guest Editor: K. N. Allen

Abstract

Biology is turning from an analytical into a synthetic discipline. This is especially apparent in the field of metabolic engineering, where the concept of synthetic metabolism has been recently developed. Compared to classical metabolic engineering efforts, synthetic metabolism aims at creating novel metabolic networks in a rational fashion from bottom-up. However, while the theoretical design of synthetic metabolic networks has made tremendous progress, the actual realization of such synthetic pathways is still lacking behind. This is mostly because of our limitations in enzyme discovery and engineering to provide the parts required to build synthetic metabolism. Here I discuss the current challenges and limitations in synthetic metabolic engineering and elucidate how modern day enzymology can help to build a synthetic metabolism of the future.

Introduction

One of the most important and disruptive events in the history of chemistry was its transformation from a purely analytical-descriptive into a synthetic-constructive discipline, which took place more than one hundred years ago [1,2]. Understanding the elemental composition of matter as well as the nature and reactivity of the chemical bond enabled chemists to use their knowledge to create new molecules and materials [3,4]. This development provided humankind with new chemical compounds, such as color dyes, pharmaceuticals, as well as polymers and plastics. Given its transforming nature, it is beyond any doubt that synthetic chemistry has been one of the key enabling technologies of the 20th century, which has virtually changed the world we are living in. Biology is currently at the verge of a similar transition [5]. Over the last decades, our ability to analyze and manipulate living systems has provided the intellectual as well as technological basis to create biological features that are new to nature.

Review

Classical metabolic engineering: Exploiting natural metabolic networks

A fundamental feature of living systems is metabolism, which can be defined as the dynamic chemistry that allows life to organize itself in three and four dimensions [6]. The incredible metabolic potential of biology is impressively demonstrated by the more than 2,000 different chemical transformations that can simultaneously take place inside of an *Escherichia coli* cell [7,8], as well as by the more than 200,000 different molecules that have been isolated from different biological systems so far [9]. This diversity has inspired generations of biologists to use living cells as small chemical factories for the production of chemicals.

In the past, many efforts centered on manipulating the metabolism of cells to obtain a target molecule. Most of these approaches were based on the concept of metabolic engineering [10]. According to this concept, known pathways and enzymes are manipulated in such a way that a certain molecule can be produced at high purity and yield from a living bacterial cell [7]. In respect to their complexity (Figure 1), these classical metabolic engineering approaches can be classified as level 1 efforts, i.e., the optimization of a natural pathway in a native organism, or level 2 efforts, i.e., the transplantation or reconstruction of a natural pathway in a new host organism [11].

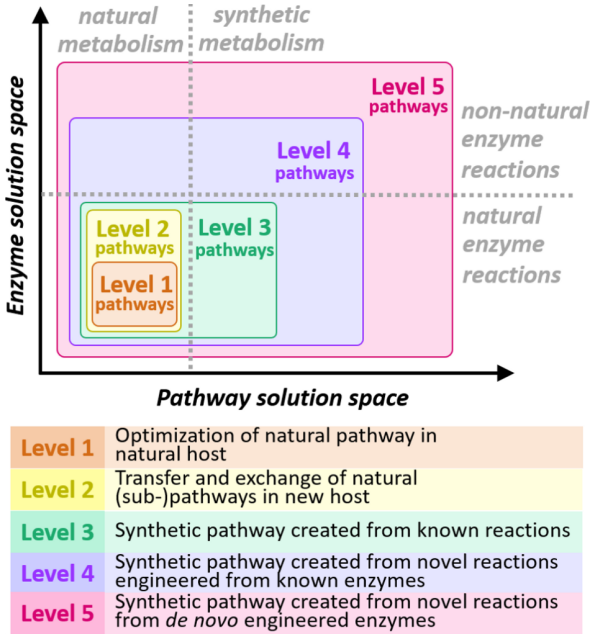


Figure 1: The five levels of metabolic engineering and their definitions according to [11]. The enzyme solution space describes the number of available enzyme reactions. The pathway solution space corresponds to the number of possible pathways that can be constructed. While level 1, 2 and 3 metabolic engineering efforts do not differ in enzyme solution space, because they all rely on known enzyme reactions, level 4 and 5 metabolic engineering efforts are built on new enzyme reactions, which expands the number of pathway solutions.

Classical metabolic engineering efforts, however, are limited in a way that they are still bound to existing pathways and reactions, which limit the accessibility of certain compounds, as well as the efficiency with which those compounds can be produced. In an ideal world, the way a target molecule is produced should not be dictated by the serendipity and constraints of evolution, but be accessible through rational design. However, this requires a fundamental understanding of those principles that are necessary for designing, realizing and operating multi-reaction sequences and metabolic networks de novo.

Metabolic retrosynthesis: Next level metabolic engineering

Recently, the concept of "synthetic metabolism" was developed that aims at overcoming the limitations provided by natural metabolism through the realization of completely novel metabolic networks [11,12]. The novel networks are designed from first principles based on simple physico-chemical considerations, such as kinetics and thermodynamics. For the design, a starting compound and a target molecule are defined and a short, thermodynamically feasible and energetically efficient route connecting the two molecules is identified. While level 3 engineering efforts aim at creating new pathway solutions by mixing and matching known enzymes from different metabolic pathways, the design efforts in their most advanced form (i.e., level 4 and level 5) do not build on existing enzymes, but only consider plausible chemical transformations and feasible metabolic intermediates [13-16]. In a subsequent realization phase, the corresponding enzymes to realize the theoretical network are identified and/or engineered and a first version of the synthetic network is reconstructed. The network is further optimized or evolved in following rounds in respect to production rate and yield.

As an example, several novel level 3 and level 4 pathways for the conversion of CO_2 into organic acids were developed recently [13,14]. These pathways are predicted to be more efficient than the naturally evolved Calvin cycle of photosynthesis, because they require less energy (ATP, redox power and/or photons) and can be supposedly operated at higher catalytic rates compared to natural carbon fixation. Accordingly, the synthetic CO_2-fixation cycles should be able to convert more carbon dioxide with less energy in a given time and hence succeed natural photosynthesis in volumetric capacity and energetic efficiency. One of these designs, the so-called CETCH cycle (Figure 2a), a synthetic level 4 pathway for the conversion of CO_2 into organic acids, was experimentally realized in vitro by combining 17 enzymes (including three engineered ones) from

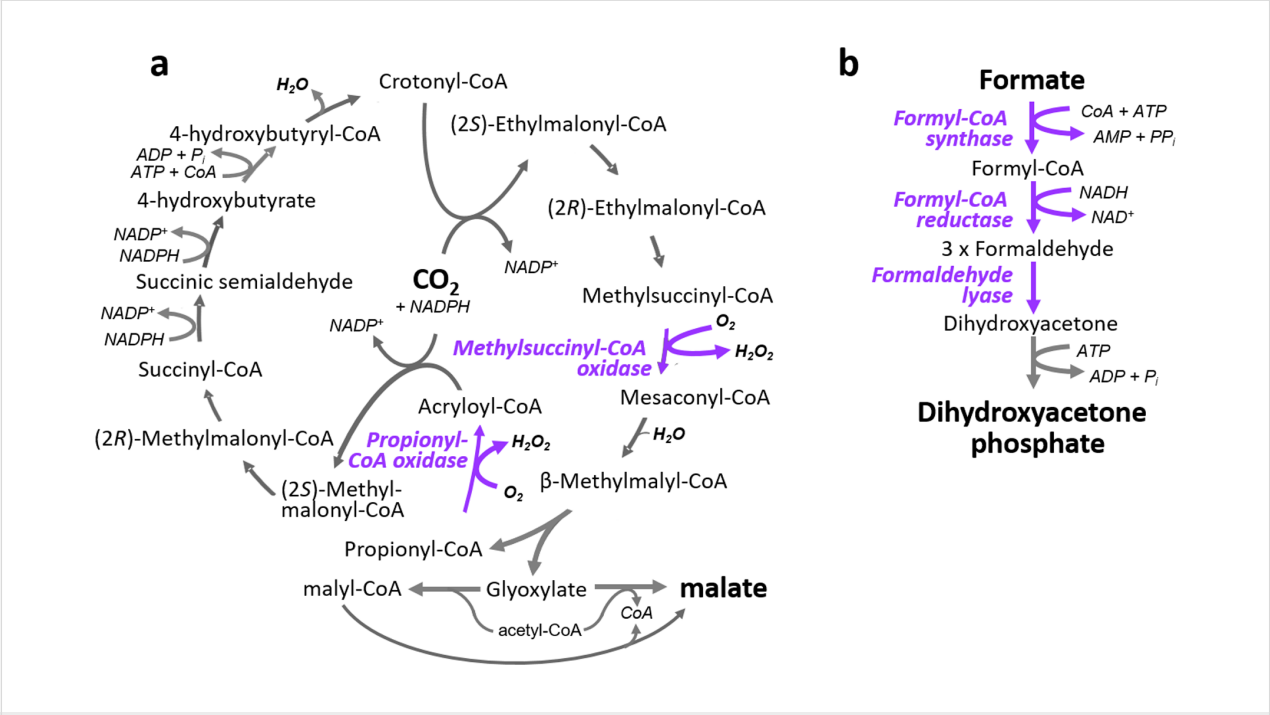


Figure 2: Two level 4 pathways that were recently realized in vitro. (a) The CETCH cycle for CO_2 fixation [13] and (b) the formolase pathway for formate assimilation [17]. Important enzymes that were engineered to establish these cycles and are mentioned in the text are highlighted in purple.

a total of nine different organisms from all three domains of life [13]. Compared to the first version of the cycle, the system was further improved until version 5.4 by almost a factor of 20, indicating that subsequent system optimization might be as important as initial reconstruction [13].

In a similar fashion, multiple level 3 and level 4 routes for the transformation of the one-carbon compound formate into cellular building blocks were designed that should theoretically outcompete natural formate assimilation pathways [18,19]. Some of the level 3 pathways were recently reconstructed in vivo [20-23] and one of the level 4 solutions – the formolase pathway (Figure 2b) – was demonstrated already in vitro [17]. This pathway relies on three new-to nature reactions, the most prominent one being the name-giving formolase reaction, which allows the subsequent condensation of three formaldehyde molecules into the three-carbon compound dihydroxyacetone phosphate [24]. In addition to that, several alternative photorespiration, methanol assimilation, as well as glycolytic pathways of levels 3 and 4 were developed that are supposedly more carbon and energy efficient compared to their naturally evolved equivalents [23,25-28].

Yet, while an increasing number of theoretical designs are proposed, the successful experimental realization of many of these designs in the lab is still falling short. This is especially true for pathways of design levels 4 and 5 that feature novel reactions, for which the corresponding enzymes are unknown (i.e., were not described to date). The realization of these pathways is severely restricted by our limited ability to discover and/or engineer new-to-nature enzymes. Notable exceptions are the formolase pathway and CETCH cycle that required each the establishment of three novel enzymatic reactions for their successful realization. However, other level 4 pathway designs require the establishment of more than ten so-far unknown enzyme reactions, emphasizing the challenge to realize truly synthetic metabolic networks [13].

The challenge of finding (new) enzymes for synthetic metabolic networks

From above examples it becomes evident that for building completely novel pathways and/or complex reaction cascades, resources are required that provide synthetic biologists with the information to find individual enzymes for a given synthetic metabolic network. More than 116 million proteins were deposited into protein sequence databases, such as UniProtKB [29]. More than 40,000 enzymes were biochemically characterized and the corresponding data is available in specialized enzyme databases, such as BRENDA [30]. This wealth of biological information provides a good starting point to search for enzyme variants that possess a desired catalytic activity.

While existing databases might provide a good resource to find the parts to reconstruct level 3 pathways, this task becomes

more challenging in respect to level 4 and level 5 designs that require new-to-nature reactions. How can these new enzyme reactions be identified or established? One option is the de novo-design of enzymes assisted by computational methods, which have been developed over the last couple of years. When combined with experimental evolution and elaborate screening methods, these efforts have allowed to establish completely novel enzyme reactions from scratch [31-34].

However, even though considerable progress has been made in creating enzymes with the help of computational methods [35], it is a complementary (and equally valid) approach to discover and/or engineer novel reactions from the natural diversity of enzymatic scaffolds [36-39]. One example is formaldehyde lyase (or "formolase") – the key enzyme of the formolase pathway – that was crafted from a benzaldehyde lyase, which showed initially some side reactivity with formaldehyde [17,24]. Other examples are propionyl-CoA oxidase and methylsuccinyl-CoA oxidase in the CETCH cycles that were engineered from a promiscuous short chain acyl-CoA oxidase and a FAD-dependent methylsuccinyl-CoA dehydrogenase, respectively [13,40,41].

These efforts in exploiting the promiscuity of enzymes to create novel catalysts might profit from new computational methods that succeeded in creating active sites of remarkable promiscuous activities in the scaffold of existing enzymes [42]. Such computationally-created "catalytically diverse active sites" could be further developed towards a new activity through directed evolution. Without any question, screening protein sequence and enzyme databases for suitable candidates is key to advance metabolic retrosynthesis. However, there are still some practical issues in extracting the necessary information from different databases. One particular problem of sequence databases like UniProtKB is the high number of misannotated proteins, which is caused by automatized annotation algorithms that are often based on "simple" sequence similarities [43,44]. In selected enzyme (super)families the annotation error can be as high as 90% [45], which masks or even impedes the identification of novel functions within a given enzyme (super)family. An example are reducing enoyl-CoA carboxylases that were for most of the time annotated as ordinary enoyl-CoA reductases, with which they are phylogenetically related [46,47]. Another example are RubisCO-like proteins [48] that are enolases [49], isomerases [50] and transcarboxylases [51], respectively, which are not capable of fixing CO_2, but are still found very often misannotated as their CO_2 fixing homologs RubisCO, with which they share a common evolutionary history [52].

A solution to overcome the problem of misannotation might come from novel computational tools that were developed recently to analyze the diversity of enzyme (super)families in respect to new functions [53-55]. While these tools have been successfully used to identify and discover new metabolic pathways (Balskus, etc.), they might as well be used to identify interesting candidate enzymes to be screened for new catalytic reactions in metabolic retrosynthesis. Further improvements in homology modeling and virtual docking are expected to increase accuracy and throughput, which will help to map and predict the substrate and reactions catalyzed by an enzyme superfamily and its individual members in the future.

Enzyme promiscuity: Key and challenge for synthetic metabolism

Another problem is that even in databases that list the experimentally confirmed activity of enzymes, an important aspect is very often not well documented: substrate (and reaction) promiscuity. Yet, this information is essential to identify suitable candidate templates to engineer or evolve a new activity within the backbone of a given enzyme. For example, although the BRENDA database is probably one of the best resources to learn about the detailed catalytic properties of enzymes, it only provides in selected cases detailed information on the activity of a given enzyme with different substrate analogs. Besides providing the necessary information to identify interesting enzyme candidates for level 4 and level 5 pathway construction, more systematic data on enzyme promiscuity would also allow a more holistic view onto the catalytic (and evolutionary) potential of a complete enzyme superfamily [56].

Note that the information on substrate and/or reaction promiscuity is not only important to establish novel enzyme reactions, it is also of very practical information in the actual construction and optimization of synthetic metabolic networks. One problem in realizing metabolic networks from scratch with enzymes that did not evolve in the same physiological context is that the individual enzymes in such mix-and-match networks are prone to feature side reactivities with substrates or products of other enzymes in the synthetic network, most likely because they lack a common evolutionary history that selects for stringent substrate specificity [57]. These unwanted side reactivities are able to compete with the wanted reactions of the synthetic network and can lead to the accumulation of dead-end products, thus decreasing or even inhibiting flux through the whole synthetic network [58]. Consequently, it is important to learn of such unwanted side reactivities before reconstruction of the network to avoid unfruitful interactions and suboptimal functioning of the system.

Again, the CETCH cycle provides a good example, why information on the promiscuity of enzymes is so important for metabolic retrosynthetic efforts. In the first versions of the synthetic

pathway, a promiscuous methylmalyl-CoA lyase caused the accumulation of malyl-CoA from an undesired side reaction of the enzyme with acetyl-CoA, which stalled the cycle. To overcome the problem of unwanted malyl-CoA accumulation, a malyl-CoA thioesterase [59] had to be added to the synthetic network. This enzyme effectively recycles the dead-end metabolite back into two intermediates of the network, malate and free CoA, thus serving as a "proof-reading" enzyme at the periphery of the CETCH cycle to keep the system running. Another problem was posed by the promiscuous activity of propionyl-CoA carboxylase with acetyl-CoA. This problem was solved by replacing the problematic reaction with another enzymatic route. Finally, an initially promiscuous acyl-CoA oxidase was further engineered to increase the catalytic efficiency for the wanted substrate propionyl-CoA compared to the unwanted substrate 4-hydroxybutyryl-CoA by a factor of 50 [13]. Having had known these problematic side reactions beforehand would have probably allowed a more rational design and/or avoided some problems upfront [58].

Yet, it needs to be mentioned that even if complex synthetic metabolic networks can be realized in vitro, this does not mean that these metabolic networks can be easily transplanted into living cells. The introduction of new reactions and metabolites into a host cell is expected to create interactions with the native metabolic and regulatory network of the host. Again, promiscuity poses a major challenge. Even though the metabolites and reactions might be completely non-native to the cell, these intermediates might be still drained due to unwanted side reactions or create unwanted metabolic and regulatory effects that the negatively affect or even prohibit operation of the synthetic metabolic network inside the host. This problem is exacerbated by the fact that for a well-studied organism like E. coli, the function of a large number of enzymes remains still unknown and there are likely to be hundreds if not thousands of unknown reactions and metabolites, often described as catalytic or metabolic "dark matter" [53,60]. Thus, a more detailed understanding of the promiscuity of native enzymes and the interaction of small molecules with the native regulatory network of cells is an important prerequisite to realize synthetic metabolism in the future [61]. In this context, it might also be very interesting to learn, which cellular hosts might be suited best for the transplantation of a given artificial network, or if current approaches to build synthetic cells from the bottom-up might represent a valuable alternative strategy [6].

Linking enzymology and synthetic biology

In summary, synthetic biology can develop its full potential, if it becomes able to harness the diversity of the millions of different enzyme variants and homologs that naturally exist. While such information is collected and made available by many enzymology and biochemistry laboratories worldwide in a community effort, it is not provided in an optimal way so that it can be used for the synthetic metabolism community. How could this apparent gap be bridged?

First, it will be necessary to collect enzymatic data in a more standardized fashion. As a matter of fact, standardization has been an important driver in the development of synthetic biology. This is probably best demonstrated by the BioBrick System [62] and the multitude of standardized genetic elements that are available for the assembly of complex genetic networks. The STRENDA standard [63,64] might provide a good blue print, how enzyme data could be organized and reported by enzymologists in the future so that the synthetic biologist could better compare and evaluate different enzymes in respect to their suitability for a given pathway.

Second it will be important not only to investigate a given enzyme in respect to its native reaction, but also study its (potential) side reactivities more systematically. For every new enzyme characterized, it would be helpful if the enzymologist tested at least a small set of substrate and/or cofactor analogs. Even though a detailed kinetic data would not necessary be required, the fact that a certain side reactivity exists in the scaffold of a given enzyme would already be a highly useful and relevant information for the synthetic biologist. On the one hand, this information could be used to identify a target enzyme for further engineering to develop the side reactivity as main activity [38,39]. On the other hand, this data would allow the synthetic biologist to anticipate potentially unwanted side reactions in the metabolic network by a given part and take corresponding countermeasures [58].

Third and lastly, there cannot enough enzymes be described. The discovery of new enzymes as well as the characterization of homologs of known enzymes needs to be continued and eventually even intensified. Only these efforts will allow to build an exhaustive library of enzyme parts for level 3, level 4 and level 5 metabolic engineering. At the same time the methods of (re-)engineering and the de-novo design of enzymes need to be further developed. This will allow to further develop and improve catalytic activities in enzymes and create new enzyme reactions that cannot be found naturally. Altogether, these activities will expand the limits of natural metabolism and pave the way for synthetic metabolic networks. Enzymology is far from being an old-fashioned business, its most fruitful era might just have begun.

Acknowledgements

The author wants to thank Carsten Kettner for organizing the 2017 Beilstein Enzymology Symposium that served as inspira-

tion for this review, as well as the two anonymous reviewers for their valuable suggestions during the review process. Research of the author on the design and realization of synthetic metabolism is supported by the Max Planck Society, the European Research Council (ERC "Syborg"), as well as the BMBF ("FormatPlant").

References

1. Jaffe, B. *Crucibles: The Story of Chemistry from Ancient Alchemy to Nuclear Fission;* Dover Publications: New York, 1976.
2. Wöhler, F. *Ann. Phys. (Berlin, Ger.)* **1828,** *88,* 253–256. doi:10.1002/andp.18280870206
3. Asimov, I. A. *Short History of Chemistry;* Anchor Books: Garden City, New York, 1967.
4. Nicolaou, K. C.; Vourloumis, D.; Winssinger, N.; Baran, P. S. *Angew. Chem., Int. Ed.* **2000,** *39,* 44–122. doi:10.1002/(sici)1521-3773(20000103)39:1<44::aid-anie44>3.0.co;2-l|
5. Yeh, B. J.; Lim, W. A. *Nat. Chem. Biol.* **2007,** *3,* 521–525. doi:10.1038/nchembio0907-521
6. Schwille, P.; Spatz, J.; Landfester, K.; Bodenschatz, E.; Herminghaus, S.; Sourjik, V.; Erb, T. J.; Bastiaens, P.; Lipowsky, R.; Hyman, A.; Dabrock, P.; Baret, J.-C.; Vidakovic-Koch, T.; Bieling, P.; Dimova, R.; Mutschler, H.; Robinson, T.; Tang, T.-Y. D.; Wegner, S.; Sundmacher, K. *Angew. Chem., Int. Ed.* **2018,** *57,* 13382–13392. doi:10.1002/anie.201802288
7. Weaver, D. S.; Keseler, I. M.; Mackie, A.; Paulsen, I. T.; Karp, P. D. *BMC Syst. Biol.* **2014,** *8,* 79. doi:10.1186/1752-0509-8-79
8. Orth, J. D.; Conrad, T. M.; Na, J.; Lerman, J. A.; Nam, H.; Feist, A. M.; Palsson, B. O. *Mol. Syst. Biol.* **2011,** *7,* 535. doi:10.1038/msb.2011.65
9. Wink, M. *Phytochemistry* **2003,** *64,* 3–19. doi:10.1016/s0031-9422(03)00300-5
10. Woolston, B. M.; Edgar, S.; Stephanopoulos, G. *Annu. Rev. Chem. Biomol. Eng.* **2013,** *4,* 259–288. doi:10.1146/annurev-chembioeng-061312-103312
11. Erb, T. J.; Jones, P. R.; Bar-Even, A. *Curr. Opin. Chem. Biol.* **2017,** *37,* 56–62. doi:10.1016/j.cbpa.2016.12.023
12. Bilgin, T.; Wagner, A. *PLoS One* **2012,** *7,* e39903. doi:10.1371/journal.pone.0039903
13. Schwander, T.; Schada von Borzyskowski, L.; Burgener, S.; Cortina, N. S.; Erb, T. J. *Science* **2016,** *354,* 900–904. doi:10.1126/science.aah5237
14. Bar-Even, A.; Noor, E.; Lewis, N. E.; Milo, R. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 8889–8894. doi:10.1073/pnas.0907176107
15. Carbonell, P.; Parutto, P.; Baudier, C.; Junot, C.; Faulon, J.-L. *ACS Synth. Biol.* **2014,** *3,* 565–577. doi:10.1021/sb4001273
16. Rodrigo, G.; Carrera, J.; Prather, K. J.; Jaramillo, A. *Bioinformatics* **2008,** *24,* 2554–2556. doi:10.1093/bioinformatics/btn471
17. Yu, H.; Liao, J. C. *Nat. Commun.* **2018,** *9,* No. 3992. doi:10.1038/s41467-018-06496-4
18. Bar-Even, A. *Biochemistry* **2016,** *55,* 3851–3863. doi:10.1021/acs.biochem.6b00495
19. Bar-Even, A.; Noor, E.; Flamholz, A.; Milo, R. *Biochim. Biophys. Acta, Bioenerg.* **2013,** *1827,* 1039–1047. doi:10.1016/j.bbabio.2012.10.013
20. Yishai, O.; Bouzon, M.; Döring, V.; Bar-Even, A. *ACS Synth. Biol.* **2018,** *7,* 2023–2028. doi:10.1021/acssynbio.8b00131
21. Bang, J.; Lee, S. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115,* E9271–E9279. doi:10.1073/pnas.1810386115
22. Tashiro, Y.; Hirano, S.; Matson, M. M.; Atsumi, S.; Kondo, A. *Metab. Eng.* **2018,** *47,* 211–218. doi:10.1016/j.ymben.2018.03.015
23. Siegel, J. B.; Smith, A. L.; Poust, S.; Wargacki, A. J.; Bar-Even, A.; Louw, C.; Shen, B. W.; Eiben, C. B.; Tran, H. M.; Noor, E.; Gallaher, J. L.; Bale, J.; Yoshikuni, Y.; Gelb, M. H.; Keasling, J. D.; Stoddard, B. L.; Lidstrom, M. E.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 3704–3709. doi:10.1073/pnas.1500545112
24. Poust, S.; Piety, J.; Bar-Even, A.; Louw, C.; Baker, D.; Keasling, J. D.; Siegel, J. B. *ChemBioChem* **2015,** *16,* 1950–1954. doi:10.1002/cbic.201500228
25. Trudeau, D. L.; Edlich-Muth, C.; Zarzycki, J.; Scheffen, M.; Goldsmith, M.; Khersonsky, O.; Avizemer, Z.; Fleishman, S. J.; Cotton, C. A. R.; Erb, T. J.; Tawfik, D. S.; Bar-Even, A. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115,* E11455–E11464. doi:10.1073/pnas.1812605115
26. Bogorad, I. W.; Lin, T.-S.; Liao, J. C. *Nature* **2013,** *502,* 693–697. doi:10.1038/nature12575
27. Erb, T. J.; Zarzycki, J. *Curr. Opin. Chem. Biol.* **2016,** *34,* 72–79. doi:10.1016/j.cbpa.2016.06.026
28. Shih, P. M.; Zarzycki, J.; Niyogi, K. K.; Kerfeld, C. A. *J. Biol. Chem.* **2014,** *289,* 9493–9500. doi:10.1074/jbc.c113.543132
29. UniProt Consortium. *Nucleic Acids Res.* **2018,** *46,* 2699. doi:10.1093/nar/gky092
30. Schomburg, I.; Jeske, L.; Ulbrich, M.; Placzek, S.; Chang, A.; Schomburg, D. *J. Biotechnol.* **2017,** *261,* 194–206. doi:10.1016/j.jbiotec.2017.04.020
31. Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. *Nat. Chem. Biol.* **2013,** *9,* 494–498. doi:10.1038/nchembio.1276
32. 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., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008,** *319,* 1387–1391. doi:10.1126/science.1152692
33. Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. *Nat. Chem.* **2017,** *9,* 50–56. doi:10.1038/nchem.2596
34. Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 10358–10363. doi:10.1073/pnas.1121063109
35. Kries, H.; Blomberg, R.; Hilvert, D. *Curr. Opin. Chem. Biol.* **2013,** *17,* 221–228. doi:10.1016/j.cbpa.2013.02.012
36. Hammer, S. C.; Kubik, G.; Watkins, E.; Huang, S.; Minges, H.; Arnold, F. H. *Science* **2017,** *358,* 215–218. doi:10.1126/science.aao1482
37. Arnold, F. H. *Angew. Chem., Int. Ed.* **2018,** *57,* 4143–4148. doi:10.1002/anie.201708408
38. Khersonsky, O.; Malitsky, S.; Rogachev, I.; Tawfik, D. S. *Biochemistry* **2011,** *50,* 2683–2690. doi:10.1021/bi101763c
39. Bornscheuer, U. T.; Kazlauskas, R. J. *Angew. Chem., Int. Ed.* **2004,** *43,* 6032–6040. doi:10.1002/anie.200460416
40. Burgener, S.; Schwander, T.; Romero, E.; Fraaije, M.; Erb, T. *Molecules* **2017,** *23,* 68. doi:10.3390/molecules23010068
41. Schwander, T.; McLean, R.; Zarzycki, J.; Erb, T. J. *J. Biol. Chem.* **2018,** *293,* 1702–1712. doi:10.1074/jbc.ra117.000764
42. Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.; Amengual-Rigo, P.; Guallar, V.; Tawfik, D. S.; Fleishman, S. J. *Mol. Cell* **2018,** *72,* 178–186.e5. doi:10.1016/j.molcel.2018.08.033

43. Zallot, R.; Harrison, K.; Kolaczkowski, B.; de Crécy-Lagard, V. *Life* **2016,** *6,* 39. doi:10.3390/life6030039
44. Radivojac, P.; Clark, W. T.; Oron, T. R.; Schnoes, A. M.; Wittkop, T.; Sokolov, A.; Graim, K.; Funk, C.; Verspoor, K.; Ben-Hur, A.; Pandey, G.; Yunes, J. M.; Talwalkar, A. S.; Repo, S.; Souza, M. L.; Piovesan, D.; Casadio, R.; Wang, Z.; Cheng, J.; Fang, H.; Gough, J.; Koskinen, P.; Törönen, P.; Nokso-Koivisto, J.; Holm, L.; Cozzetto, D.; Buchan, D. W. A.; Bryson, K.; Jones, D. T.; Limaye, B.; Inamdar, H.; Datta, A.; Manjari, S. K.; Joshi, R.; Chitale, M.; Kihara, D.; Lisewski, A. M.; Erdin, S.; Venner, E.; Lichtarge, O.; Rentzsch, R.; Yang, H.; Romero, A. E.; Bhat, P.; Paccanaro, A.; Hamp, T.; Kaßner, R.; Seemayer, S.; Vicedo, E.; Schaefer, C.; Achten, D.; Auer, F.; Boehm, A.; Braun, T.; Hecht, M.; Heron, M.; Hönigschmid, P.; Hopf, T. A.; Kaufmann, S.; Kiening, M.; Krompass, D.; Landerer, C.; Mahlich, Y.; Roos, M.; Björne, J.; Salakoski, T.; Wong, A.; Shatkay, H.; Gatzmann, F.; Sommer, I.; Wass, M. N.; Sternberg, M. J. E.; Škunca, N.; Supek, F.; Bošnjak, M.; Panov, P.; Džeroski, S.; Šmuc, T.; Kourmpetis, Y. A. I.; van Dijk, A. D. J.; ter Braak, C. J. F.; Zhou, Y.; Gong, Q.; Dong, X.; Tian, W.; Falda, M.; Fontana, P.; Lavezzo, E.; Di Camillo, B.; Toppo, S.; Lan, L.; Djuric, N.; Guo, Y.; Vucetic, S.; Bairoch, A.; Linial, M.; Babbitt, P. C.; Brenner, S. E.; Orengo, C.; Rost, B.; Mooney, S. D.; Friedberg, I. *Nat. Methods* **2013,** *10,* 221–227. doi:10.1038/nmeth.2340
45. Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. *PLoS Comput. Biol.* **2009,** *5,* e1000605. doi:10.1371/journal.pcbi.1000605
46. Erb, T. J.; Berg, I. A.; Brecht, V.; Muller, M.; Fuchs, G.; Alber, B. E. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 10631–10636. doi:10.1073/pnas.0702791104
47. Erb, T. J.; Brecht, V.; Fuchs, G.; Muller, M.; Alber, B. E. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106,* 8871–8876. doi:10.1073/pnas.0903939106
48. Hanson, T. E.; Tabita, F. R. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98,* 4397–4402. doi:10.1073/pnas.081610398
49. Ashida, H.; Saito, Y.; Kojima, C.; Kobayashi, K.; Ogasawara, N.; Yokota, A. *Science* **2003,** *302,* 286–290. doi:10.1126/science.1086997
50. Erb, T. J.; Evans, B. S.; Cho, K.; Warlick, B. P.; Sriram, J.; Wood, B. M.; Imker, H. J.; Sweedler, J. V.; Tabita, F. R.; Gerlt, J. A. *Nat. Chem. Biol.* **2012,** *8,* 926–932. doi:10.1038/nchembio.1087
51. Carter, M. S.; Zhang, X.; Huang, H.; Bouvier, J. T.; Francisco, B. S.; Vetting, M. W.; Al-Obaidi, N.; Bonanno, J. B.; Ghosh, A.; Zallot, R. G.; Andersen, H. M.; Almo, S. C.; Gerlt, J. A. *Nat. Chem. Biol.* **2018,** *14,* 696–705. doi:10.1038/s41589-018-0067-7
52. Erb, T. J.; Zarzycki, J. *Curr. Opin. Biotechnol.* **2018,** *49,* 100–107. doi:10.1016/j.copbio.2017.07.017
53. Zallot, R.; Oberg, N. O.; Gerlt, J. A. *Curr. Opin. Chem. Biol.* **2018,** *47,* 77–85. doi:10.1016/j.cbpa.2018.09.009
54. Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. *PLoS One* **2009,** *4,* e4345. doi:10.1371/journal.pone.0004345
55. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. *Biochim. Biophys. Acta, Proteins Proteomics* **2015,** *1854,* 1019–1037. doi:10.1016/j.bbapap.2015.04.015
56. Vögeli, B.; Erb, T. J. *Curr. Opin. Chem. Biol.* **2018,** *47,* 94–100. doi:10.1016/j.cbpa.2018.09.013
57. Barelier, S.; Cummings, J. A.; Rauwerdink, A. M.; Hitchcock, D. S.; Farelli, J. D.; Almo, S. C.; Raushel, F. M.; Allen, K. N.; Shoichet, B. K. *J. Am. Chem. Soc.* **2014,** *136,* 7374–7382. doi:10.1021/ja501354q
58. Sun, J.; Jeffryes, J. G.; Henry, C. S.; Bruner, S. D.; Hanson, A. D. *Metab. Eng.* **2017,** *44,* 150–159. doi:10.1016/j.ymben.2017.10.006
59. Erb, T. J.; Frerichs-Revermann, L.; Fuchs, G.; Alber, B. E. *J. Bacteriol.* **2010,** *192,* 1249–1258. doi:10.1128/jb.01267-09
60. Ellens, K. W.; Christian, N.; Singh, C.; Satagopam, V. P.; May, P.; Linster, C. L. *Nucleic Acids Res.* **2017,** *45,* 11495–11514. doi:10.1093/nar/gkx937
61. Donati, S.; Sander, T.; Link, H. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **2018,** *10,* e1396. doi:10.1002/wsbm.1396
62. Shetty, R. P.; Endy, D.; Knight, T. F., Jr. *J. Biol. Eng.* **2008,** *2,* 5. doi:10.1186/1754-1611-2-5
63. Gardossi, L.; Poulsen, P. B.; Ballesteros, A.; Hult, K.; Švedas, V. K.; Vasić-Rački, Đ.; Carrea, G.; Magnusson, A.; Schmid, A.; Wohlgemuth, R.; Halling, P. J. *Trends Biotechnol.* **2010,** *28,* 171–180. doi:10.1016/j.tibtech.2010.01.001
64. Swainston, N.; Baici, A.; Bakker, B. M.; Cornish-Bowden, A.; Fitzpatrick, P. F.; Halling, P.; Leyh, T. S.; O'Donovan, C.; Raushel, F. M.; Reschel, U.; Rohwer, J. M.; Schnell, S.; Schomburg, D.; Tipton, K. F.; Tsai, M.-D.; Westerhoff, H. V.; Wittig, U.; Wohlgemuth, R.; Kettner, C. *FEBS J.* **2018,** *285,* 2193–2204. doi:10.1111/febs.14427

License and Terms