Biochemical and Biophysical Characterization of TyrA Enzymes from Symbiotic Hyperthermophilic Archaea *Nanoarchaeum equitans* and *Ignicoccus hospitalis*

A Ph.D. Thesis

In the Department of Chemistry and Biochemistry

Presented by Irina Shlaifer

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

April 2016

This is to certify that the thesis prepared

By: Irina Shlaifer

Entitled: Biochemical and Biophysical Characterization of TyrA Enzymes from Symbiotic Hyperthermophilic Archaea *Nanoarchaeum equitans* and *Ignicoccus hospitalis*

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy Chemistry (Biochemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair

Dr. Natalie Philips

_____ External Examiner

Dr. Susan Aitken

_____ External to Program

Dr. David Kwan

_____ Examiner

Dr. Paul Joyce

_____ Examiner

Dr. Peter Pawelek

_____ Thesis Supervisor

Dr. Joanne Turnbull

Approved by: _____

Dr. Heidi Muchall, Graduate Program Director

April 2016 _____André Roy, Dean of Faculty

Abstract

The biosynthesis of L-tyrosine (L-Tyr) and L-phenylalanine (L-Phe) is directed by the interplay of three enzymes. Chorismate mutase (CM) catalyzes the rearrangement of chorismate to prephenate, which can be either converted to hydroxyphenylpyruvate by prephenate dehydrogenase (PD) or to phenylpyruvate by prephenate dehydratase (PDT). This work reports the first characterization of both the trifunctional PD-CM-PDT from the smallest hyperthermophilc archaeon (*Nanoarchaeum equitans*) and the bifunctional CM-PD from its host (the crenarchaeon *Ignicoccus hospitalis*). Hexa histidine-tagged proteins were expressed in *Escherichia coli* and purified by chromatography on Ni-NTA affinity resin. Both enzymes were highly thermally stable and exhibited maximal activity at 90°C. CM, PD and PDT activities were detected at temperatures consistent with enzymes from extreme thermophiles. Kinetic analysis revealed that unlike most PDs, the two archaeal enzymes were insensitive to regulation by L-Tyr and preferred $NADP^+$ to NAD^+ as a cofactor in the dehydrogenase reaction. *N. equitans* PDT was feedback inhibited by L-Phe (K_i = 0.8 μM) in a non-competitive fashion consistent with L-Phe's combination at a site separate from that of prephenate. Gel filtration and analytical ultracentrifugation analysis of bifunctional CM-PD from *I. hospitalis* suggested that the enzyme is a native dimer. Limited proteolysis studies revealed that the enzyme is highly resistant to proteolysis but could be cleaved to yield a stable C-terminal PD domain. Mass spectrometry and mutagenesis studies confirmed that the PD domain of bifunctional *I. hospitalis* CM-PD could be independently isolated and expressed. Biochemical and biophysical characterization of this active truncated variant was performed and the results of solution studies were compared to those of the full-length protein and to information available from other PD enzymes. Guided by amino acid sequence alignment predictions and

by models based on the available crystal structures of bacterial homologues, eight variants containing site-specific replacements were generated in *I. hospitalis* CM-PD as attempts to alter cofactor selectivity and substrate and end-product binding. Those variant proteins were kinetically characterized in order to help define the role of active site residues in substrate/ inhibitor interactions. These are the first studies exploring the aromatic amino acid biosynthetic pathway from the two archaeal organisms, which provide efficient and stable catalysts as excellent candidates for applications in biotechnology.

Acknowledgements

I would like to thank my supervisor Dr. Joanne Turnbull for her direction and guidance, my committee members Dr. Joyce and Dr. Pawelek for their scientific advice, Dr. Pamela Hanic-Joyce and Dr. M. J. Kornblatt for their help and Dr. Ann English, Dr. Dajana Vuckovic, Alain Tessier and Dr. Heng Jiang for mass spectrometry assistance. I also acknowledge my colleagues Dr. Matthew Leibovitch and Dr. Francis MacManus for their help with modelling. I would also like to thank my family for their outstanding support, encouragement and patience. My loving husband Daniel, my parents: Marina and Sergey Sapozhnikov, my sister Hanna (Anna), my parents-in-law: Anastasia and Gennady, and of course my children: David and Victoria.

Table of Contents

List of Figures

List of Tables

List of abbreviations

Å	angstroms
ACN	acetonitrile
AD	arogenate dehydrogenase
AUC	analytical ultracentrifugation
CD	circular dichroism
CM	chorismate mutase
CM-PD	chorismate mutase–prephenate dehydrogenase
CM-PDT	chorismate mutase–prephenate dehydratase
Da	Dalton
DL-FTyr	m-fluoro-D, L-tyrosine
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
ESI-MS	electrospray ionization mass spectrometry
FA	formic acid
FPLC	fast protein liquid chromatography
FPLC-SEC	size exclusion FPLC
Gdn-HCl	guanidinium hydrochloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HPP	(4-hydroxyphenyl)pyruvate

IhTyrA	TyrA protein from *I. hospitalis*
IPTG	isopropyl-β-D-thiogalactopyranoside
K_M	Michaelis constant
L-Phe	L-phenylalanine
L-Tyr	L-tyrosine
MS	mass spectrometry
MW	molecular weight
NAD(P)	oxidized form of nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	reduced form of nicotinamide adenine dinucleotide (phosphate)
NeTyrA	TyrA protein from *N. equitans*
Ni-NTA	nickel-nitrilotriacetic acid
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PD	prephenate dehydrogenase
PDT	prephenate dehydratase
PP	phenylpyruvate
TFA	trifluoroacetic acid
Tris	tris(hydroxymethyl)aminomethane
SDS	sodium dodecylsulfate
U	units
V_{max}	maximum velocity

Chapter 1

General Introduction

1.0 Aromatic amino acid biosynthesis

The aromatic amino acids, L-tyrosine (L-Tyr), L-tryptophan (L-Trp) and L-phenylalanine (L-Phe) are required for the growth and survival of all living organisms. Not only do these amino acids serve as the building blocks of proteins but they are also precursors in the synthesis of essential aromatic metabolites such as flavonoids[(1)], quinones[(2)], cyanogenic glycosides[(3)] and alkaloids[(4)]. In archae, bacteria, plants, fungi and apicomplexan parasites the biosynthesis of these aromatic amino acids is a multistep enzymatic process which involves the "shikimate pathway" and the "common pathway"; in contrast mammals obtain these amino acids from their diet[(5)]. Thus, enzymes that are involved in these pathways are attractive targets for the design of inhibitors which potentially can serve as antifungal and antibacterial agents, and as herbicides[(6-8)]. Moreover, these enzymes are of great interest in protein engineering since the aromatic amino acids are precursors for commercially valuable products such as food supplements, melanin, biodegradable polymers, and the anti-Parkinson drug L-Dopa and its derivatives [(9-14)]. Accordingly, a thorough understanding of the mechanism of action and mode of regulation of these enzymes is essential for the design of effective inhibitors as well as for the overproduction of the aromatic amino acids through metabolic engineering.

1.1 Shikimate pathway

The shikimate pathway links carbohydrate metabolism to the biosynthesis of aromatic compounds (Fig 1.1) [(15)]. It consists of seven metabolic steps that convert derivatives of six-carbon sugars such as glucose into chorismate[(16)]. The pathway commences with the condensation of erythrose-4-phosphate and phosphoenol-pyruvate to yield a seven-carbon compound 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP). The reaction is catalyzed by the highly regulated DAHP synthase, whose isoforms are feedback inhibited by either L-Phe, L-

Tyr or L-Trp[16]. In the following steps, DAHP is cyclized to yield shikimate which is then phosphorylated to shikimate 3-phosphate and converted to 5-enolpyruvylshikimate-3-phosphate (EPSP). In the last step of the shikimate pathway, EPSP is converted to chorismate by chorismate synthase (Fig 1.1). The shikimate pathway was lost in eukaryotes through evolution and therefore is absent from humans. As such, considerable efforts have been made to design herbicides that are not toxic to humans and to develop drugs against diseases such as malaria and tuberculosis by targeting enzymes in the pathway[6]. One of the best known effective herbicides is glyphosate (Roundup®), which inhibits 5-enolpyruvylshikimate 3-phosphate synthase[17, 18].

1.2 Common pathway

Chorismate is a key branch-point intermediate that serves as a common precursor for the synthesis of the three aromatic amino acids L-Phe, L-Tyr and L-Trp as well as for a number of aromatic compounds such as vitamins, quinones or folates (Fig 1.2) [19, 20]. In the common pathway chorismate undergoes a Claisen rearrangement to prephenate, the common intermediate for the biosynthesis of L-Tyr and L-Phe. This reaction is catalyzed by chorismate mutase (CM) and is considered unique since it is the only known example of an enzyme-catalyzed Claisen rearrangement in nature. The remaining biosynthetic steps leading to the formation L-Tyr and L-Phe can occur by two possible routes: the hydroxyphenylpyruvate/ phenylpyruvate (HPP/PP) route or the arogenate route (Fig 1.2). The biosynthesis of the third amino acid, L-Trp, also originates from chorismate and contains six steps from a separate pathway. The four reactions of this pathway are catalyzed by enzyme complexes, namely the anthranilate synthase-phosphoribosyl transferase complex and the tryptophan synthase

complex. The L-Trp biosynthetic pathway has gained industrial interest and has been exploited through metabolic engineering for the production of aromatic compounds such as bio-indigo[21-23].

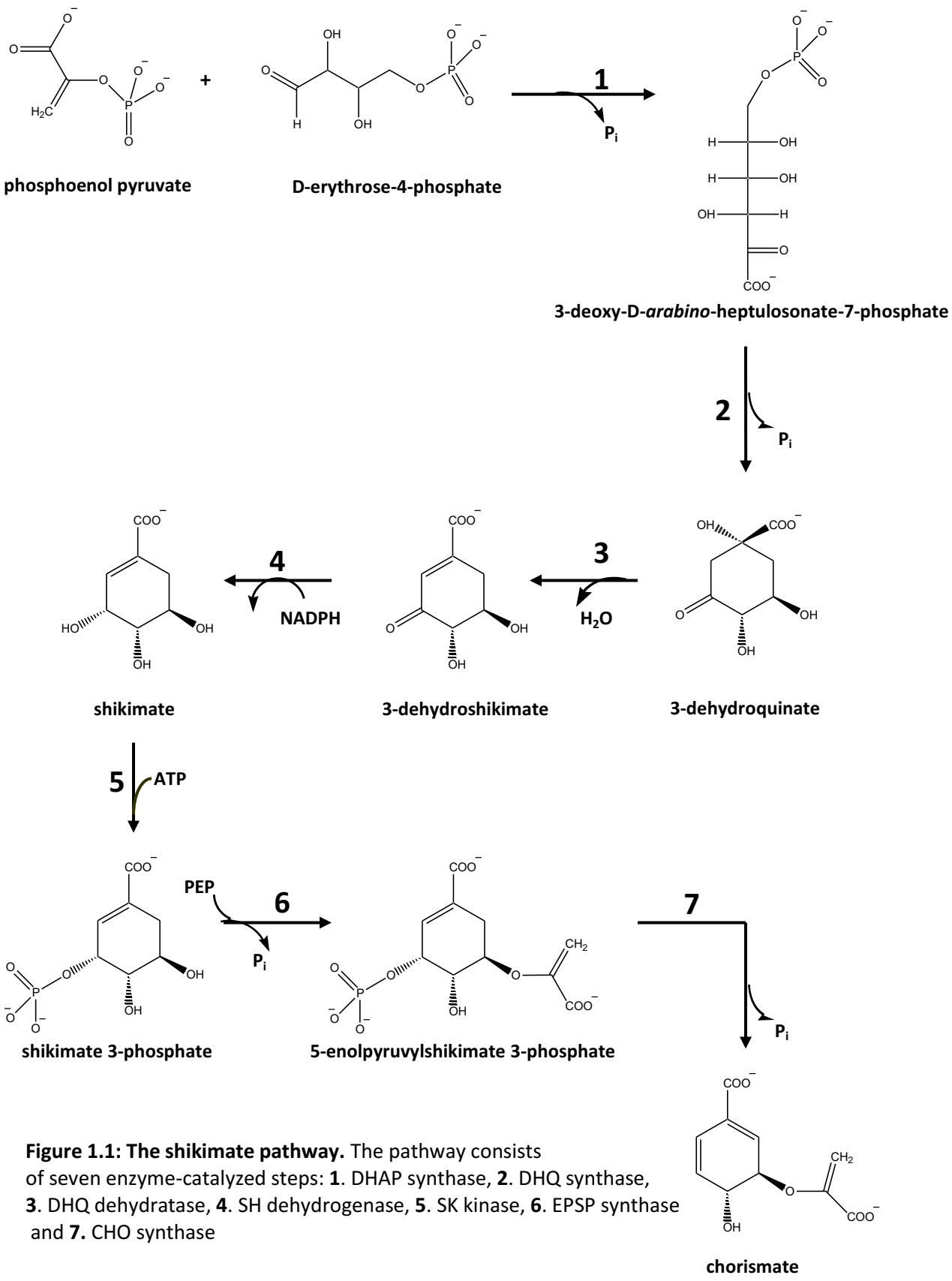


Figure 1.1: The shikimate pathway. The pathway consists of seven enzyme-catalyzed steps: **1**. DHAP synthase, **2**. DHQ synthase, **3**. DHQ dehydratase, **4**. SH dehydrogenase, **5**. SK kinase, **6**. EPSP synthase and **7.** CHO synthase

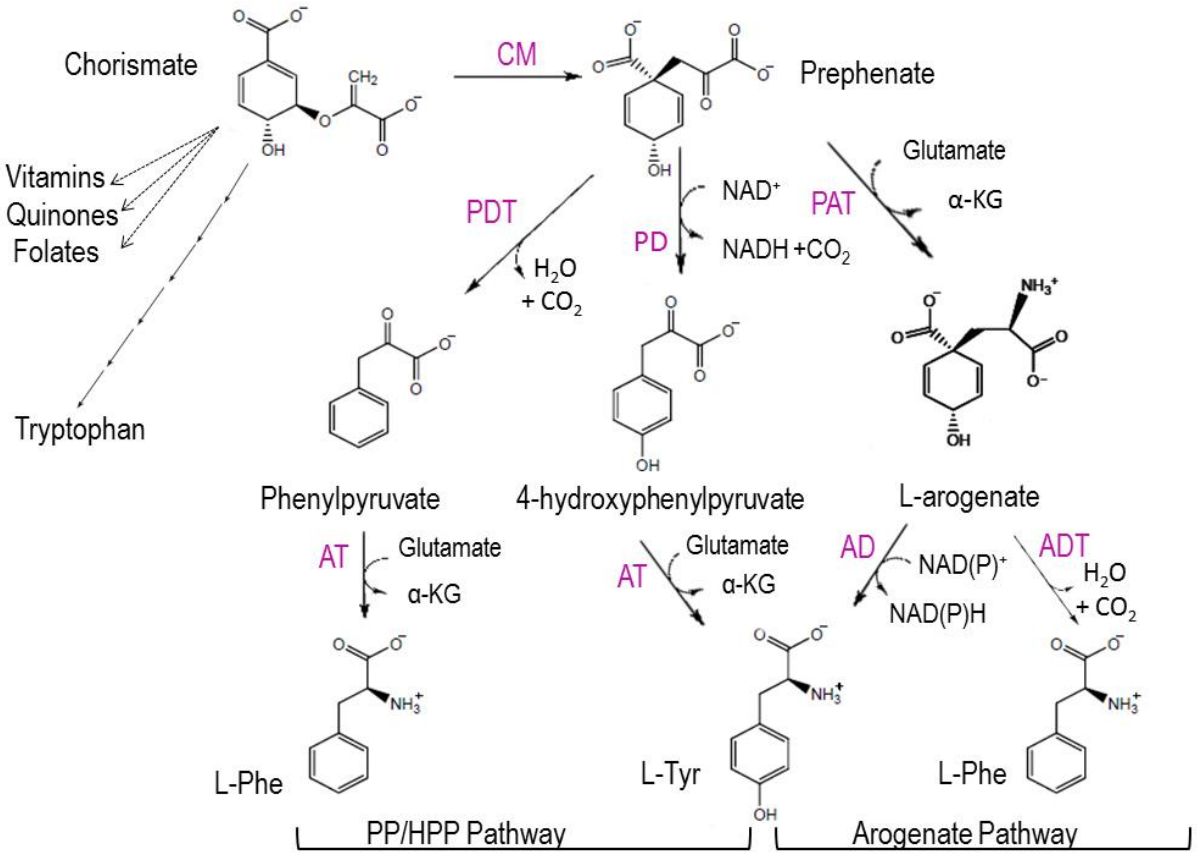


Figure 1.2: Biosynthesis of L-tyrosine and L-phenylalanine. Biosynthesis of the aromatic amino acids can occur by the phenylpyruvate/4-hydroxyphenylpyruvate (PP/HPP) pathway and the arogenate pathway. Enzymes involved in the biosynthesis: CM - chorismate mutase, PD – prephenate dehydrogenase, PDT – pephenate dehydratase, AT – aminotransferase, AD- arogenate dehydrogenase, ADT – arogenate dehydratase.

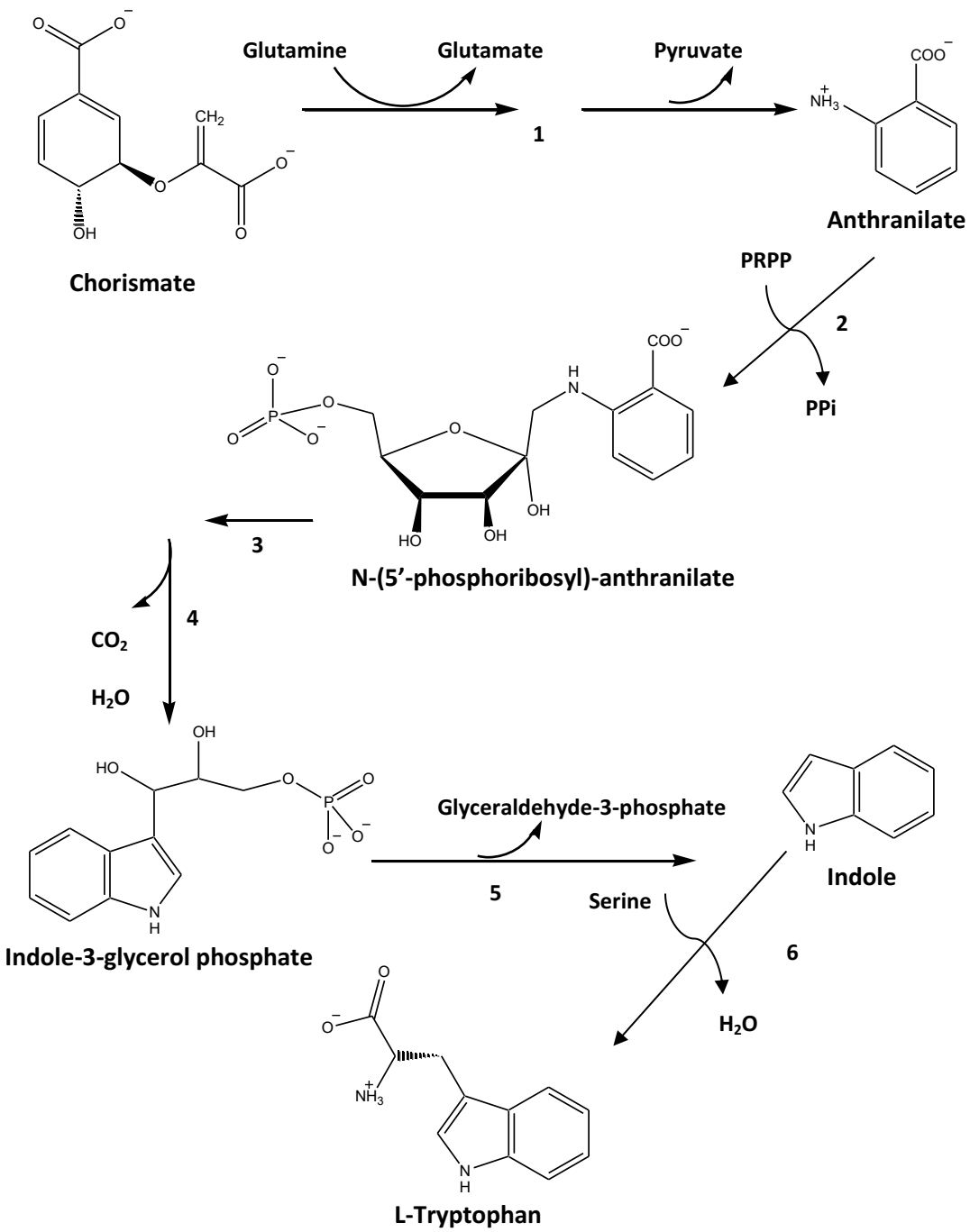


Figure 1.3: L-Tryptophan biosynthesis. The pathway consists of six enzyme-catalyzed steps: **1**, anthranilate synthase, **2**, anthranilate phosphoribosyltransferase, **3**, phosphoribosylanthranilate isomerase, **4**, indole-3-glycerol-phosphate synthase, **5**-**6**, tryptophan synthase enzyme complex. PRPP: 5-phosphoribosyl-Δ-pyrophosphate (Adapted from J. Bonvin [(24)]).

1.3 The hydroxyphenylpyruvate and phenylpyruvate pathways

In bacteria, the production of L-Tyr and L-Phe generally occurs through the HPP/PP route where prephenate, the product of the chorismate mutase (CM) reaction, is a branch point intermediate. Prephenate can either undergo $NAD(P)^+$- dependent oxidative decarboxylation by prephenate dehydrogenase (PD) to form *p*-hydroxyphenylpyruvate (HPP) and carbon dioxide, or dehydration and decarboxylation by prephenate dehydratase (PDT) to yield phenylpyruvate (PP) (Fig 1.2). HPP or PP are then transaminated by an aminotransferase to form L-Tyr or L-Phe, respectively[25]. In many organisms the HPP pathway is regulated by the end product L-Tyr which inhibits the action of PD and to a lesser extent CM[26]. Similarly, the PP route is also regulated by its end product, L-Phe, which allosterically inhibits PDT. L-Phe is also reported to inhibit mutase activity but the dehydratase enzyme is significantly more affected than the mutase[27].

1.4 The arogenate pathway

Most plants and some cyanobacteria produce L-Tyr and L-Phe using the arogenate pathway[28, 29]. In this pathway prephenate is first transaminated into L-arogenate by prephenate aminotransferase (PAT) using either L-glutamate or L-aspartate as amino donors (Fig 1.2). L-arogenate (formerly denoted as pretyrosine) is then either oxidatively decarboxylated by arogenate dehydrogenase (AD) in the presence of NAD^+ or $NADP^+$ to form L-Tyr and carbon dioxide, or dehydrated by arogenate dehydratase (ADT) to form L-Phe. AD has been isolated from various species and was shown in plants to be very sensitive to feedback inhibition by Tyr[30, 31]. In contrast, ADT also was reported to be feedback regulated by L-Phe in species such as rice, tobacco and spinach [32, 33] and stimulated by L-Tyr[34].

While most organisms possess either the HPP/PP or the arogenate pathway for L-Tyr and L-Phe production, in some microorganisms such as *Zymomonas mobilis, Pseudomonas*

aeruginosa and *Synechocystis* both pathways co-exist [35-41]. In these organisms L-arogenate is used for L-Tyr synthesis while L-Phe is synthesized via the PP route. In these cases L-arogenate is not at the branch-point and therefore AD is not feedback inhibited by the end product L-Tyr.

Biosynthetic routes to produce L-Tyr and L-Phe are of great interest in bioengineering since these amino acids and their derivatives have commercial value in the food, pharmaceutical and agricultural industries [42, 43]. An excellent example of commercial success in the food industry is the sweetener aspartame, also known as NutraSweet® which consist of L-Phe (and L-aspartate)[44, 45]. Traditionally, L-amino acids have been supplied mainly by extraction from protein hydrolysates[46]. However, to suit the demands for large scale production of the L-α-keto acids and intermediates, strategies to manipulate their biosynthetic pathways are now being explored [12, 47, 48]. Accordingly, characterization of the key enzymes in these pathways has been performed in a number of organisms. In this thesis, enzymes responsible for L-Tyr and L-Phe production namely CM, PD and PDT are examined to provide insights regarding their structure, activity, and the relationship between the enzymes which carry out the consecutive reactions.

1.5 TyrA protein family

The TyrA protein family is dedicated to L-Tyr biosynthesis and consists of homologous cofactor-dependent dehydrogenases that are classified into three categories depending on their substrate specificities: prephenate dehydrogenases (PD) are specific to prephenate, arogenate dehydrogenases (AD) to L-arogenate, and cyclohexadienyl dehydrogenases that can accept both substrates. In addition to specificity for the cyclohexadienyl substrate, TyrA enzymes have a requirement for either NAD^+ or $NADP^+$, while some may use both cofactors[29]. Generally, prephenate dehydrogenases are specific to NAD^+, for example *E. coli* and *Aquifex aeolicus* PDs, while arogenate dehydrogenases prefer $NADP^+$, for example *Synechocystis* sp. PCC 6803 AD [41].

It has been proposed that broad cofactor specificity is the ancestral state prior to the divergence of the NAD^+ and $NADP^+$-specific descendants which evolved in response to the mechanisms that enhanced the cellular abundance of each of the cofactors [29, 49]. Examples for the TyrA proteins that can accept both cofactors include prephenate-specific enzymes from the Gram-negative bacterium *Gluconobacter oxydans* and from the archaeon *Methanohalophilus mahii* [29, 50]. All TyrA proteins catalyze the irreversible oxidative step in L-Tyr biosynthesis, regardless of the organism in which they are found. They share a core catalytic domain of about 30 kDa and maintain the same frame of fundamental reaction chemistry [51]. The TyrA family consists of prephenate dehydrogenases that are monofunctional, such as PD from the hyperthermophilic bacterium *Aquifex aeolicus*[52], those that are bifunctional with chorismate mutase activity associated with the N-terminal portion of the protein, for example *E. coli* and *Haemophilus influenzae* CM-PD[26], or bifunctional with phosphoshikimate carboxyvinyltransferases activity at the proteins' C-terminal region, such as PDs from *Pseudomonas stutzeri* and *Pseudomonas aeruginosa*[51]. Trifunctional PD proteins have now been identified in the genomes of *Archaeoglobus fulgidus, Nanoarchaeum equitans* and in the brown algae *Ectocarpus siliculosus* [46, 53-55]. In these organisms PD, CM and PDT domains are predicted to reside on the same polypeptide chain. Some TyrA proteins, for example PD from *Bacillus subtilis*, also possess a carboxy-terminal ACT fusion domain named after the enzymes where the domain was discovered (Aspartate kinase-Chorismate mutase-TyrA). The ACT domain is reported to be responsible for allosteric regulation of the catalytic activities [56].

Many enzymes within the TyrA family have been extensively studied in terms of their phylogeny and have been classified by substrate specificity and mode of regulation through bioinformatics analysis [29]. However, relatively few Tyr proteins have been purified and characterized. The most well-characterized TyrA enzyme to date is the bifunctional CM-PD from

E. coli, which has been extensively studied in solution using kinetic, biochemical and biophysical tools and serves as a model for the study of other TyrA proteins[26, 57-60]. However, no crystal structure is available for *E. coli* CM-PD. Other characterized TyrA proteins include the monofunctional PD from *A. aeolicus* [52], ADs from *Synechocystis* sp. PCC6803[49] and *Arabidopsis thaliana*[31] and cyclohexadienyl dehydrogenases from Z*ymomonas mobilis*[40].

1.6 Chorismate mutase mechanism

CM catalyzes the pericyclic Claisen rearrangement reaction of chorismate to prephenate[61]. The reaction can occur in the absence of the enzyme although the process is accelerated by over a million-fold in the presence of CM [62, 63]. Chorismate is found in two forms in aqueous solution: the more abundant diequatorial and the less stable diaxial form[57]. Both the non-enzymatic[64, 65] and enzyme catalyzed rearrangements[62, 66, 67] are believed to proceed via a chair-like transition state following the selection of chorismate's less stable diaxial form. It's enolpyruvyl group is stabilized by hydrogen bonding to water molecules, which leads to bond breakage between C-5 and the oxygen of chorismate (Fig 1.4).

In order to identify features that are required for enzymatic catalysis, several chorismate analogues had been synthesized and characterized [68, 69]. These studies have suggested that the allyl vinyl ether and the two carboxylate groups are required for binding of chorismate to the mutase active site while neither the 5,6-olefinic ring nor the 4-hydroxyl group are necessary[69]. An *endo*-oxabicyclic diacid inhibitor [70] possessing a bridged ether oxygen and an *endo* conformation of the bridged carboxylate (Fig 1.4) was shown to mimic the bicyclic structure of the transition state. As revealed from studies on *E. coli* CM-PD, this transition state analogue appeared to bind about 300-fold more tightly to the enzyme than chorismate[26].

Additionally, kinetic studies on *E. coli* CM-PD have revealed that the reaction proceeds through enzymic acids and bases[57, 71].

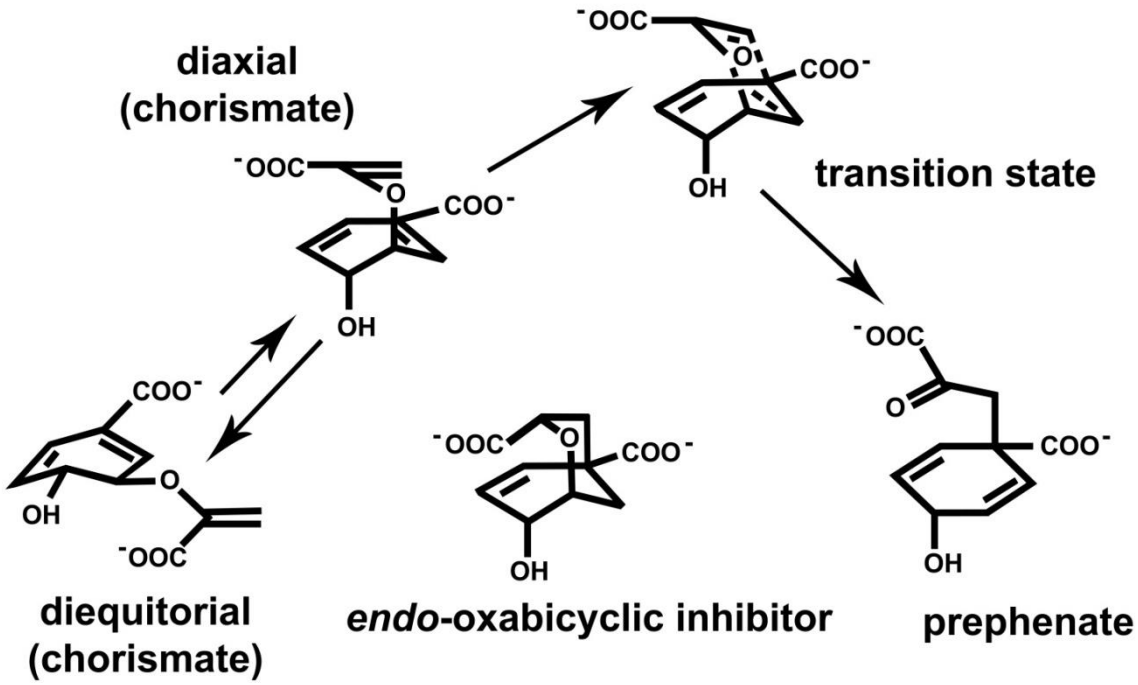


Figure 1.4: Rearrangement of chorismate through a transition-state complex. Proton NMR studies indicate that 10 to 20% of chorismate is found in the diaxial conformer in equilibrium with the more stable diequatorial form. The *endo*-oxabicyclic CM inhibitor is believed to mimic the structure of the transition state. Adapted from Christendat *et al*[57].

Several CM enzymes such as those from *Bacillus subtilis*[64], *Saccharomyces cervisiae*[72], *Thermus thermophilus*[73], *Mycobacterium tuberculosis*[74] and the independently expressed CM domain of *E. coli* CM-PDT[75] (also called the "mini-mutase") have been crystallized. The studies have revealed two structurally distinct classes of CM: AroH and AroQ (Fig 1.5). AroH is less abundant and possesses a trimeric α/β-barrel structure [20]. The monofunctional CM from *B. subtilis* which was the first to be crystallized is of the AroH type. In contrast, AroQ is

a dimeric helix-bundle, which in some organisms is fused to a PD or PDT domain. The *E. coli* mini-mutase is AroQ type. Other AroQ examples include the CM from *S. cerevisiae* and the CM from *M. tuberculosis*. Interestingly, there are two types of AroQ proteins. Those from *Pseudomonas aeroginusa* and *Salmonella thepidum*, for example, contain a cytoplasmic fusion of CM-PD or CM-PDT as well as a periplasmic monofunctional CM[(76)]. Although alignment of the amino acid sequences of the CM enzymes mentioned above shows little similarity and the proteins adopt different folds, the electronic environment and the geometry of the active site appears to be well conserved.

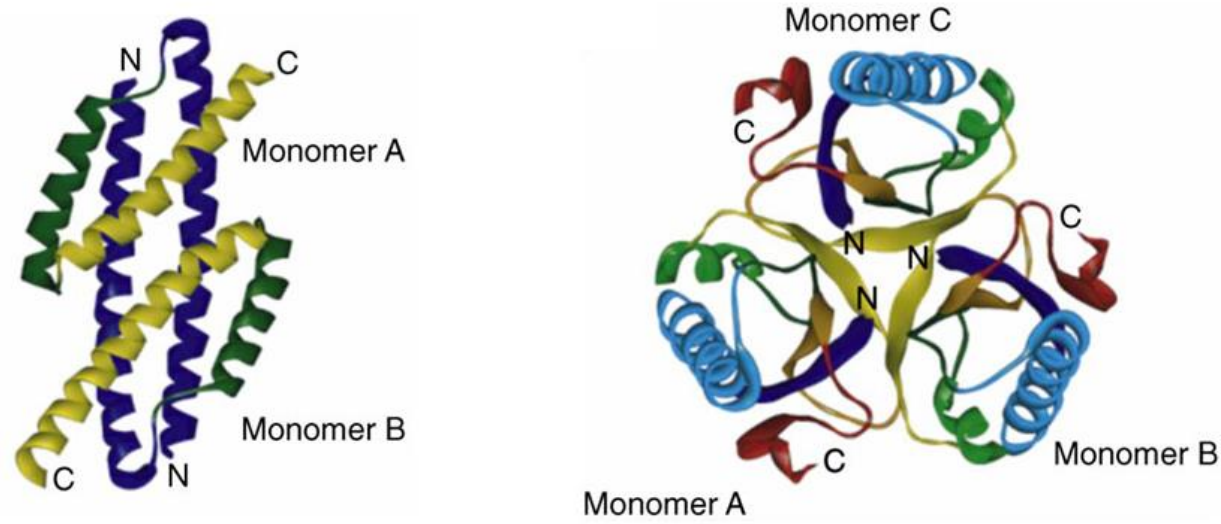
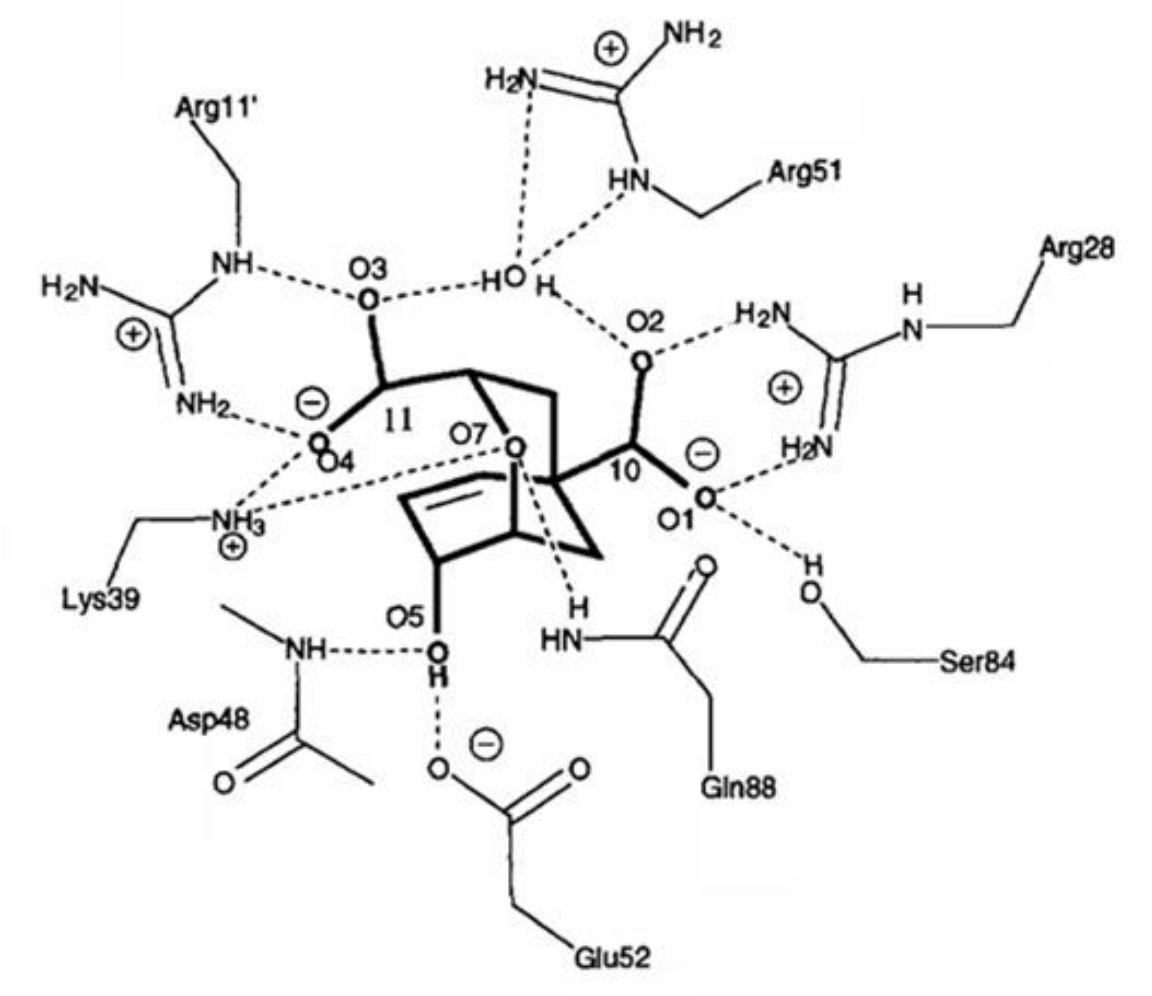


Figure 1.5: Ribbon diagram representations of the AroQ and AroH folds. The AroQ class (left) is completely helical and includes CM proteins from *E. coli* and *S. cerevisiae*. The AroH class (right) is organized as a trimeric α/β-barrel fold and includes CM proteins from *B. subtilis* and *M. tuberculosis*[(61, 77)].

Crystallography studies on the CMs have provided insights as to the active site structure required for substrate binding and catalysis [64, 72, 73, 75, 78, 79]. These findings have served as a basis for the design of complementary site-directed mutagenesis experiments and for directed evolution studies [80, 81]. The structure of the "mini-mutase" complexed with a mutase transition-state analogue[75], along with extensive mutagenesis studies[76, 77] have identified the importance of Lys39 and Gln88 in coordinating the ether oxygen (O7) of the *endo*-oxabicyclic diacid (Fig 1.6). The studies also have shown that Lys39 (Lys37 in *E. coli* CM-PD) along with Arg11 from the adjacent monomer are important in positioning the substrate's C-11 carboxylate group in the highly charged region of the active site (Fig 1.6). Kinetic studies on *E. coli* CM-PD have revealed that the activity of the mutase is pH dependent indicating the participation of three groups (two protonated and one deprotonated) in substrate binding and/or catalysis; Lys37 may be one of these residues, poised to protonate the ether oxygen of chorismate in the transition state of the reaction[71]. Additionally, pH-dependent activity profiles of the variant Gln88Glu of the "mini-mutase" and of wild-type CM from yeast (which contains Glu at position 246) have revealed the importance of a protonated side chain at this position. CM's reactivity has been recently explored through computational studies to complement the findings obtained from functional analysis of variant proteins[82-84]. Although a number of mechanisms have been proposed for the CM-catalyzed reaction which include acid/base catalysis, a nucleophile-assisted dissociative mechanism[65] and transition state stabilization, the precise catalytic mechanism is not yet completely understood and remains under active investigation.



1.6: The active site of *E. coli* chorismate mutase. Scheme of the active site of the independently expressed mutase domain ("mini-mutase") of *E. coli* CM-PDT with *endo*-oxabicyclic diacid. Residue Arg11' is from the adjacent monomer. Adapted from Lee and colleagues[61].

1.7 Prephenate dehydrogenase mechanism and its inhibition by tyrosine

Prephenate dehydrogenase catalyzes the $NAD(P)^+$-dependent oxidative decarboxylation of prephenate to HPP. The reaction is irreversible, driven by the formation of the aromatic product, HPP. In the absence of the enzyme, but only under acidic conditions, prephenate can rapidly undergo decarboxylation to yield phenylpyruvate (PP) [(85)]. The acid-induced decarboxylation follows a stepwise mechanism where protonation of the hydroxyl group of prephenate leads to the formation of a resonance-stabilized carbonium ion prior to decarboxylation. In contrast, the PD-catalyzed reaction proceeds through concerted hydride transfer and decarboxylation steps. This mechanism was postulated by Hermes and colleagues through the use of isotope effect studies performed with prephenate and its analogues [(85)]. These studies reported an isotope effect for the hydride transfer to NAD^+ when using the substrate analogue deoxoprephenate, deuterated at C-4. Using the natural abundance of ^{13}C in the substrate, they observed a carbon isotope effect for the cleavage of C-C bond between the cyclohexadiene ring and the ring carboxylate in the decarboxylation step. Most importantly however, the ^{13}C isotope effect was smaller when studies were conducted in the presence of the deuterated versus non deuterated substrate suggesting that hydride transfer and decarboxylation occurred in the same transition state. These results supported the idea of a concerted mechanism in catalysis.

Studies on *E. coli* CM-PD, which included the analyses of initial velocity, product and dead-end inhibition patterns, and isotope trapping experiments, were used to elucidate the kinetic mechanism of the dehydrogenase reaction. These studies revealed that PD conforms to a rapid equilibrium random kinetic mechanism with catalysis as the rate-determining step in the reaction[(86)]. A similar reaction mechanism has been reported for CM-PD from *A. aerogenes*[(87)]. In

addition, pH rate studies on *E. coli* CM-PD have been performed to determine the pK_a values of residues involved in the dehydrogenase reaction[59, 85, 88] . The pH dependence of the log *V* of the PD-catalyzed reaction identified a single ionizing group with a pK_a 6.5 that had to be deprotonated for maximum activity and was involved in catalysis and/or product release[59]. In contrast, the $\log(V/K)_{prephenate}$ pH profile displayed, in addition to the deprotonated group, a second ionizing group (pK_a value of about 8.4), which must be protonated for maximum activity and was important for binding prephenate to the enzyme-NAD^+ complex. Additionally, studies on the effects of temperature and solvent on the acidic pK_a limb of the pH profile suggested that the catalytic group was likely a histidine[85] .

Studies performed by Christendat and Turnbull which include site-directed mutagenesis and the analysis of pH-rate profiles have identified residues involved in substrate binding and catalysis (Fig 1.7). His197 was shown to play a key role in catalysis of the PD reaction likely by polarizing the *p*-hydroxyl group of prephenate, lowering the activation barrier to assist in hydride transfer and concominant decarboxylation; the activity of the His197Asn variant was reduced by five orders of magnitude relatively to the wild-type enzyme and retained full CM activity[57]. In contrast, Arg294 was shown to be important solely in prephenate binding as the K_M for prephenate of the Arg294Gln variant was increased over 100-fold without affecting turnover of the PD-catalyzed reaction or the CM reaction [58]. Inhibition studies using substrate analogues modified at the C-1 position of prephenate, suggested that Arg294 interacts electrostatically with the ring carboxylate of prephenate (Fig 1.7). Multiple sequence alignment with a number of PD proteins shows these residues are conserved (Fig 1.8). Interestingly, mutagenesis studies so far have failed to identify the residue (pK_a of 8.4) titrating in the pH rate profiles that is believed to assist in prephenate binding.

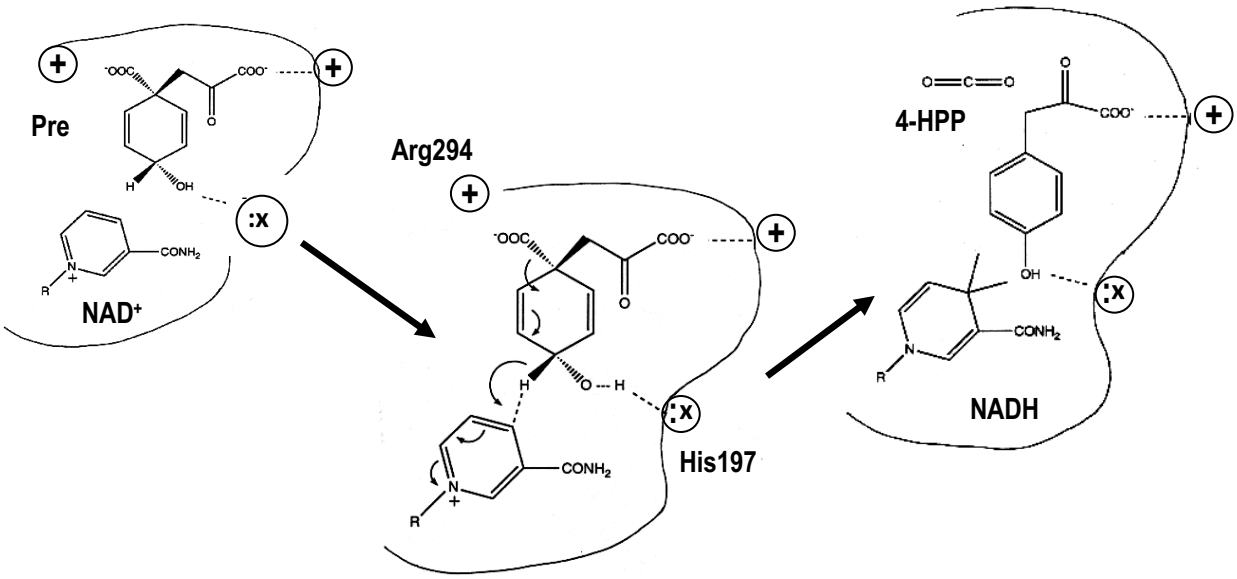


Figure 1.7: Proposed mechanism for the prephenate dehydrogenase-catalyzed reaction. A deprotonated group, His197 in *E. coli* CM-PD, is proposed to polarize the 4-hydroxy group of prephenate while Arg294 is involved in prephenate binding.

Figure 1.8: Multiple sequence alignment of TyrA proteins. Multiple sequence allignment of selected prephenate dehydrogenases and PD domains of TyrA proteins. Shown are sequences from monofunctional PDs: *Aquifex aeolicus* NP_214202, *Corynebacterium glutamicum* NP_599479.1 and *Sinorizobium meliloti* WP_010970109.1; monofunctional AD: *Synechocystis* P73906_SYNY3; bifunctional CM-PDs: *Ignicoccus hospitalis* YP_00143548, *Escherichia coli* NP_289153 and *Haemophilus influenzae* YP_005829274; and trifunctional PD-CM-PDTs: *Nanoarchaeum equitans* NP_963486.1and *Archaeoglobus fulgidus* NP_069065. Conserved residues highlighted in black correspond to the glycine residues from the GXGXXG motif characteristic of the $NAD(P)^+$ binding domain, the catalytic histidine and the arginine residue proposed to be important for prephenate binding. Shown in bold are aspartate and arginine residues that are important for $NAD(P)^+$ specificity, plus substituted residues at those two positions.

Crystallographic data for TyrA proteins have surfaced only in the last decade. The first structure was that of a monofunctional PD (the N-terminally truncated variant, Δ19PD) from the hyperthermophilic bacterium *A. aeolicus* in complex with NAD^+ that was solved by Christendat,Turnbull and coworkers[89]. This was followed by studies on the same enzyme in the presence of NAD^+ and a series of prephenate analogs including L-Tyr[90]. Published structures are now avalible for NAD^+-bound PD from *S. mutants*[91], $NADP^+$-bound AD from *Synechocystis* sp. PCC 6803 [41] and for the N-terminally truncated variant (Δ80) of CM-PD from *H. influenzae* in complex with NAD^+ and L-Tyr[92]. Unpublished crystal structures include those of $NADP^+$ -bound TyrA from *Sinorhizobium meliloti* (pdb-4wji) and unliganded PDs from *Streptococcus thermophilus* (pdb-3dzb) and *Corynebacterium glutamicum* (pdb-3ktd). These studies reveal that the dehydrogenases from all organisms mentioned above retain the same higher order structures. The enzyme is dimeric and contains within each monomer a highly α-helical C-terminal dimerization domain and an N-terminal nucleotide binding domain containing a β-α-β repeating structure that is associated with the Rossman fold [93]. The active site (one for each monomer) is at the interface of the two domains and encompasses amino acid residues from both monomers (See Fig 1.9 A).

The structures of *A. aeolicus* Δ19PD and that of the independently expressed PD domain of *H. influenzae* CM-PD (the latter shares 60% amino acid sequence identity with *E. coli* PD) have been instrumental in helping to assign the role of residues in the catalytic mechanism of the dehydrogenase. In fact, *H. influenzae* Δ80CM-PD in complex with NAD^+ and L-Tyr has served as a template to generate the liganded model of *E. coli* PD [94]. This template indicated that the cationic guanidine group of residue Arg297, which is equivalent to Arg294 in *E. coli* CM-PD, is within hydrogen bonding distance of the side chain carboxyl group of L-Tyr (Fig 1.9). Site-directed mutagenesis studies on the Δ80CM-PD from *H. influenzae* have revealed that the K_M of

the variant Arg297Gln was increased 1000-fold relative to the wild-type enzyme, denoting the importance of the arginine in prephenate binding[95]. The liganded model of *E. coli* PD also showed that the *p*-hydroxyl group of L-Tyr is located near His197 which further supports histidine's role as a catalytic residue in the PD reaction.

Crystallographic studies on *A. aeolicus* Δ19PD[89, 90] revealed that residue His147, which is equivalent to His197 in *E. coli* CM-PD, is located adjacent to the cofactor's nicotinamide ring and the C-4 position of HPP thus implicating this residue in hydride transfer from prephenate to NAD^+ (Fig 1.10). Additionally, the residue Arg250 (Arg294 in *E. coli* CM-PD), appears to be located in a highly polar environment and in close proximity to the pyruvyl side chain of HPP, implying its importance in prephenate binding through the side chain carboxylate. Site-directed mutagenesis and kinetic studies, however, revealed that Arg250 is not critical for prephenate binding in *A. aeolicus Δ19*PD[96]; the variant Arg250Gln displayed only a 10-fold increase in the K_M for prephenate without a significant change in k_{cat}. These studies identified another active site residue, Lys246' from the adjacent monomer, which in combination with Arg250, may play an important role in binding prephenate.

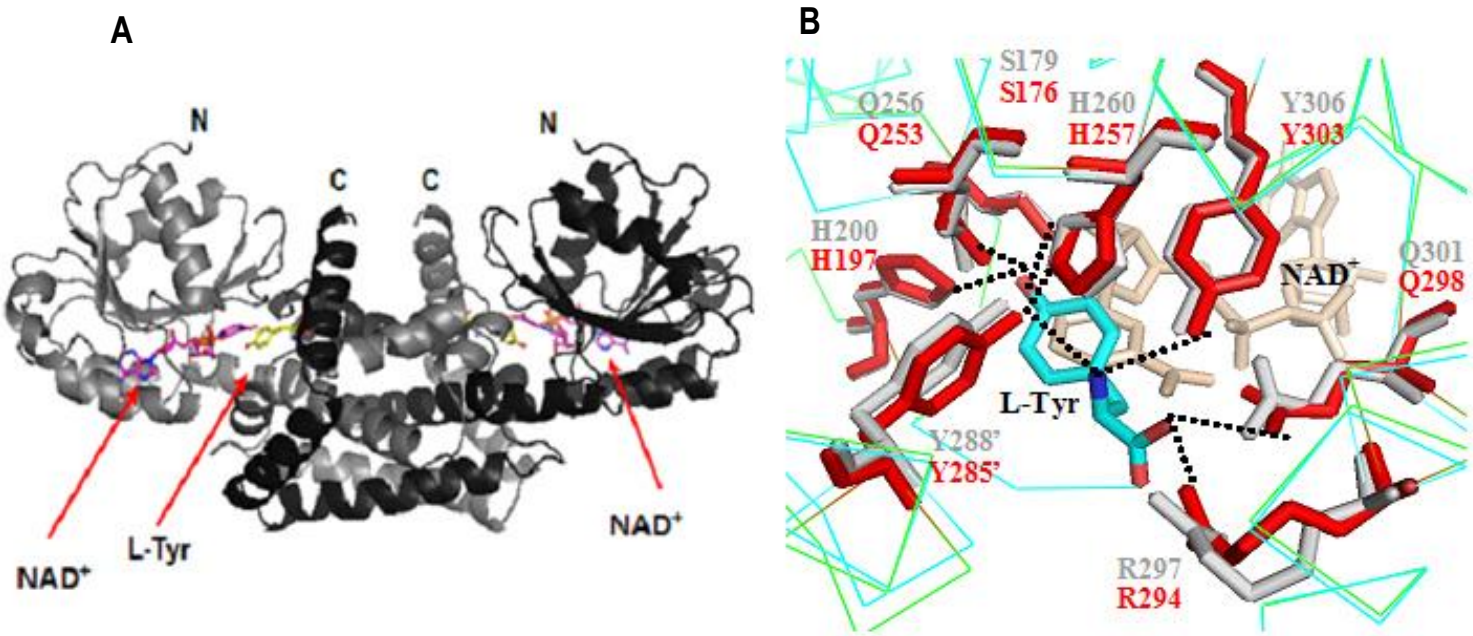


Figure 1.9: Model of *E. coli* PD domain. (A) Model of dimeric PD domain of *E. coli* CM-PD in complex with NAD^+ and L-Tyr. Based on *H. influenzae* 3D Structure: PDB-2pv7 [92, 94]. (B) Active site of liganded dimeric PD domain of *E. coli* CM-PD overlayed on the structure of *H. influenzae* PD[94]. *E. coli* PD residues are colored in red while those of *H. influenzae* PD are in white.

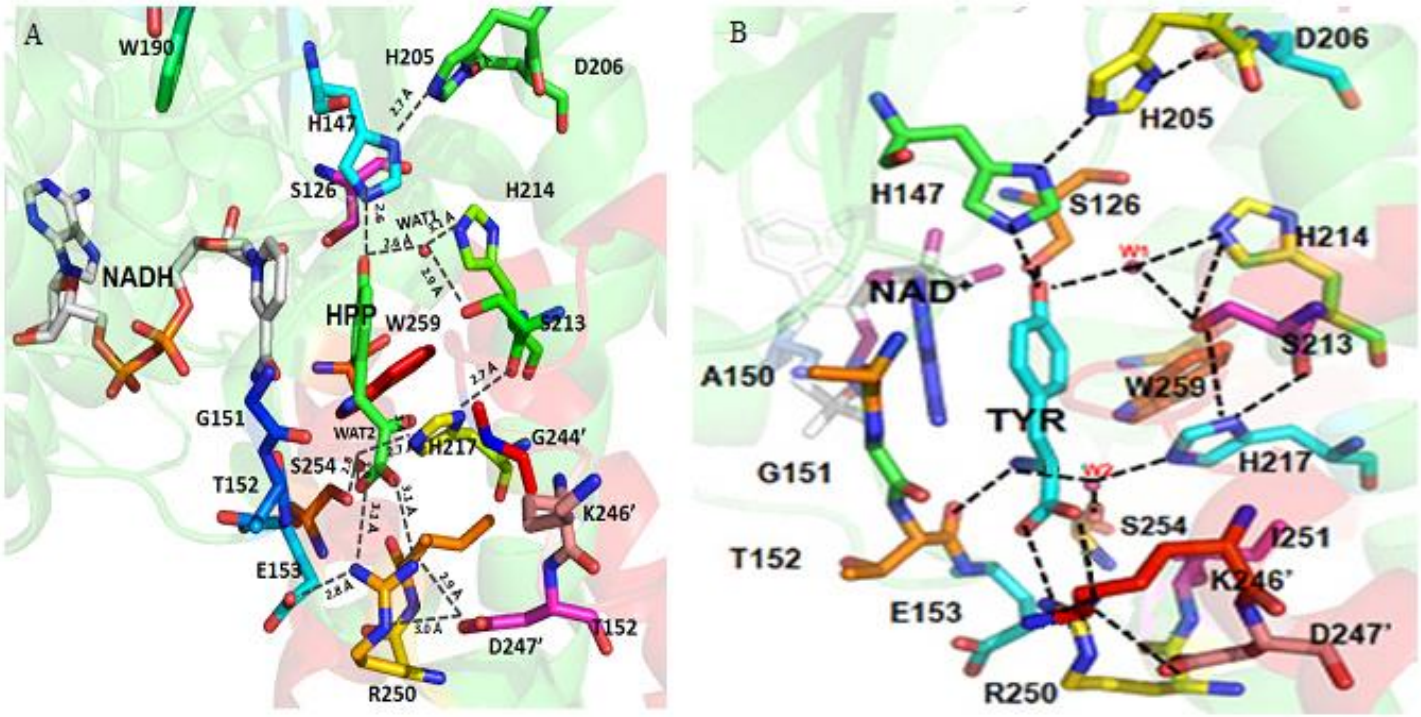


Figure 1.10: Active site of *A. aeolicus* Δ19PD. The protein is bound in (A) with HPP and NADH and in (B) with NAD^+ and L-Tyr**.** Adapted from N. Hotz [96].

L-Tyr biosynthesis is mainly regulated through feedback inhibition of TyrA proteins by the end products; however the exact regulatory mechanism is still not well understood. Kinetic studies on *E. coli* CM-PD by Christopherson and colleagues showed that L-Tyr and HPP act as competitive inhibitors[97] suggesting that the product and the end product inhibitor combine at the same site. In contrast, Turnbull and coworkers, through fitting models of inhibition, postulated the presence of a distinct allosteric site for the binding of L-Tyr[71]. The *E. coli* enzyme showed concave upwards double reciprocal plots of velocity at varying prephenate concentrations, and fixed, increasing concentrations of L-Tyr suggesting positive cooperativity between subunits in the binding of the end product [60]. Biophysical studies support a model for this allosteric transition, which involves conversion from an active dimer to an inactive tetramer upon binding L-Tyr in the presence of NAD^+ [98]. Studies on *A. aeolicus* PD and *H. influenzae* CM-PD have also displayed kinetics suggesting cooperative binding of L-Tyr, however, this did not appear to correlate with a quaternary structure change [52, 95]. Interestingly, despite the evidence in support of an allosteric binding site for L-Tyr [71, 96], the crystal structures of *H. influenzae* Δ80CM-PD and *A. aeolicus* Δ19PD clearly show L-Tyr bound at the active site (Fig 1.9-1.10 and references[90, 92]. The structure of *H. influenzae* Δ80CM-PD and the modeled PD domain of *E. coli* CM-PD as well as site-directed mutagenesis studies of *E. coli* CM-PD have identified active site residues Tyr285' and Tyr303 (each from the adjacent monomer), and Gln298 as key players in the enzyme's interactions with the amine and carboxylate side chain of L-Tyr, respectively [94]. *A. aeolicus* Δ19PD, in contrast, does not contain the coordinating enzymic tyrosine residues implying an alternative mechanism for L-Tyr binding (Fig 1.10). The structure of *A. aeolicus* Δ19PD showed L-Tyr's amine interacting with His217 via a water molecule and the backbone of Thr152 (Fig 1.10)[90]. Mutagenesis studies have also shown that His217 and Ser254 are important for inhibition by L-Tyr. Additionally, the variants Thr152Pro and

Glu153Ala/Asp247Ala are partially L-Tyr insensitive and yield kinetics consistent with allosteric inhibition by L-Tyr[96].

Additionally, bioinformatics analysis and phylogenetic studies by Song *et al* [99] have revealed that some bacterial PD proteins, such as the monofunctional PD from *B. subtilis* possess a C-terminal ACT fusion domain that is important for allosteric regulation of catalytic activity by L-Tyr and L-Phe[100]. This domain is not present in the TyrA proteins of *E. coli, H. influenzae* and *A. aeolicus*. Together the results of many studies imply that different mechanisms of inhibition by L-Tyr exist for PD proteins from different bacterial sources.

1.8 Prephenate dehydratases

Prephenate dehydratase (PDT) belongs to the PheA family of proteins that catalyze decarboxylation and dehydration of prephenate to yield phenylpyruvate (Fig 1.2). This PDT-catalyzed conversion is the committed step in the biosynthesis of L-phenylalanine in the shikimate pathway [25, 101]. In the absence of enzyme, the acid-catalyzed reaction proceeds in two steps via a carbocation intermediate (Fig 1.11) [85]. Kinetic studies, however, reveal that PDT accelerates this reaction by a factor of $>10^6$ [62, 102]. Although a number of PDT enzymes have been characterized biochemically [103-107], their catalytic mechanism remains poorly understood.

In different organisms PDTs exist as either monofunctional or multifunctional enzymes, usually as a fusion protein with CM. The *E. coli* PheA, for example, is a bifunctional enzyme that has a PDT domain fused to the N-terminus of CM (CM-PDT) [102]. PDT proteins contain a C-terminal regulatory ACT domain which binds L-Phe or other amino acids [107]. Binding of L-Phe generally has an inhibitory effect, whereas tyrosine often activates the enzyme. Regulation of PDT activity by tryptophan, methionine and leucine has also been reported[108, 109]. To elucidate important structural and functional features required for catalysis and allosteric regulation,

bioinformatics studies were conducted by Jensen and colleagues which identified nine highly conserved amino acids[110]. Of these, site-directed mutagenesis studies on PDTs from *E. coli* and *Corynebacterium glutamicum* have identified a conserved TRF motif (Thr, Arg, Phe) that is essential for dehydratase activity[103, 111]. The structural and functional importance of this motif was revealed through crystal structures of PDT from *Chlostridium tepidum* in complex with L-Phe and unliganded PDT from *Staphylococcus aureus*[112]. These studies revealed that PDT was a tetramer formed from a dimer of dimers with the active site located in a cleft between two dimeric subunits (Fig 1.12). Within this cleft the side chains of Thr171 and Phe173 (*C. tepidum* numbering) were pointing toward the active site. From isotope effect studies on PDT from *Methanocaldococcus jannaschii,* Hilvert and coworkers proposed that in the PDT reaction (illustrated in Fig 1.11) Thr acts as a general acid by protonating the leaving hydroxyl group of prephenate, while Phe forms part of a hydrophobic pocket to help anchor prephenate in the active site [113].

The crystal structure of PDT from *C. tepidum* showed L-Phe bound at the dimer interface of the ACT domains of the protein[112], confirming the allosteric regulation of PDT. These studies showed Phe242, which appears to be the critical residue for L-Phe binding, interacting with the benzene ring of L-Phe[105].

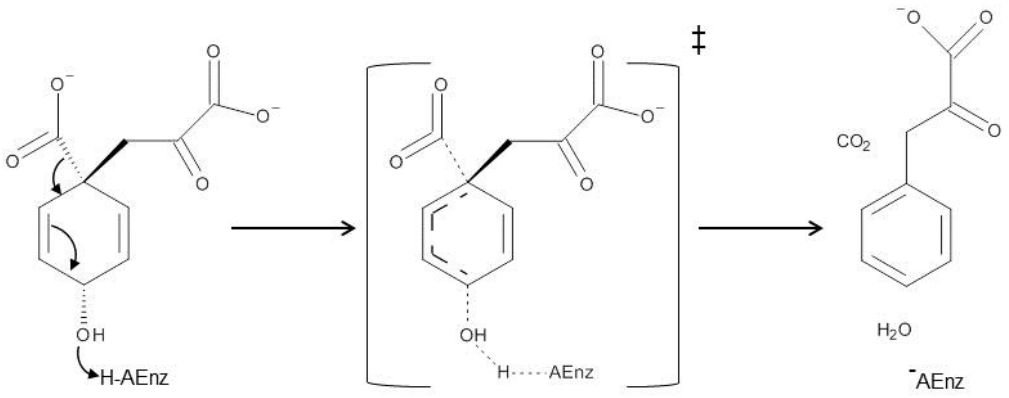


Figure 1.11: PDT-catalyzed reaction. Adapted from Van Vleet and coworkers[(113)]

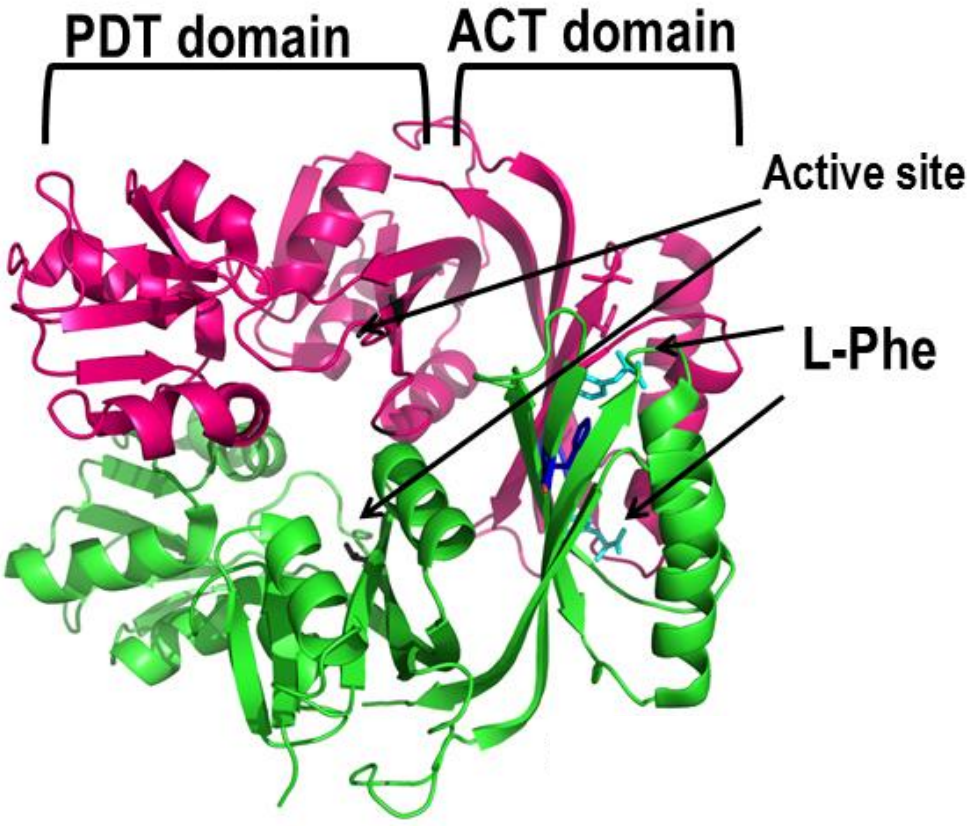


Figure 1.12: (A) Crystal structure of dimeric *Chlostridium tepidum* PDT in complex with L-Phe. The structure is from PDB-2qmx.The active site is located in the PDT domain in a cleft between the two subunits[(112)].

1.9 Arogenate dehydrogenases and dehydratases

In most plants, cyanobacteria, algae and several other microorganisms, both L-Tyr and L-Phe are synthesized from a common precursor L-arogenate via the arogenate pathway. This pathway was discovered by Stenmark *et al*[28] who named the precursor pretyrosine since they observed that this compound was used for L-Tyr biosynthesis. Later it was shown that pretyrosine could also be used in L-Phe biosynthesis and therefore was renamed L-arogenate[114]. As mentioned in section 1.4, prephenate aminotransferase (PAT) first converts prephenate to L-arogenate. Next, arogenate dehydrogenase (AD) catalyzes the cofactor-dependent oxidative decarboxylation of L-arogenate to yield L-Tyr, while arogenate dehydratase (ADT) catalyzes decarboxylation and dehydration to produce L-Phe. In contrast to PD proteins, ADs often have a strict requirement for $NADP^+$ but in some cases, such as the AD from *Synechococcus* sp. ATCC 29404, is reported to use both as cofactors[29]. Amino acid sequence alignments of AD and PD proteins by Bonner and colleagues have revealed insight into cofactor specificity[29]. The presence of an Asp or Glu residue approximately 15-20 residues downstream of the GXGXXG consensus motif is an indicator for NAD^+ specificity since $NADP^+$ is repelled by the negative charge of Asp. The absence of Asp or Glu at the corresponding position in the AD protein and the presence of a positively charged Arg at the following position indicates $NADP^+$ specificity[29].

Many studies on AD [30, 31, 38, 39, 99, 115-117] have focused on the identification of this activity in different organisms through bioinformatics analysis combined with functional studies; only a few studies have been conducted on purified protein. The most recent work is focused on feedback regulation of AD in plants[118]. In addition, studies on ADT proteins, which are much less common than PDTs have shown feedback regulation by L-Phe[32, 33] and stimulation by L-

Tyr[34]. As with PDT, the regulation of ADT is thought to be allosterically controlled through the ACT domain.

The only available AD crystal structure is from *Synechocystis* sp. PCC 6803 AD in complex with its cofactor $NADP^+$ and with the substrate, L-arogenate, modeled in the active site. It was published in 2006 by Legrand and coworkers[41]. Examination of the crystal structure revealed that *Synechocystis* sp. PCC 6803 AD is similar to the PD proteins with regard to its higher order structure and the active site configuration. Analysis of the AD active site identified the conserved residue His112 (His197 in *E. coli* CM-PD and His147 in *A. aeolicus* Δ19PD), presumed to be the reaction's catalytic hydrogen bond acceptor, in close proximity to the modeled L-arogenate (2.6 Å) [41] implying a similar catalytic mechanism to that of PD (Fig 1.13). However, the conserved residue Arg217 in AD from *Synechocystis* sp. PCC 6803 (homologous to Arg294 and Arg250 in *E. coli* CM-PD and *A. aeolicus* Δ19PD, respectively), which was shown to be important for prephenate binding in *E. coli* and *H. influenzae* CM-PDs[58, 95], appeared too far away from the active site to interact with L-arogenate implying that this cationic residue was not important for substrate binding [41]. To date, there are no published crystal structures of any ADT proteins. More studies, thus, are required to provide insights into the structure-function relationship of ADs and ADT proteins.

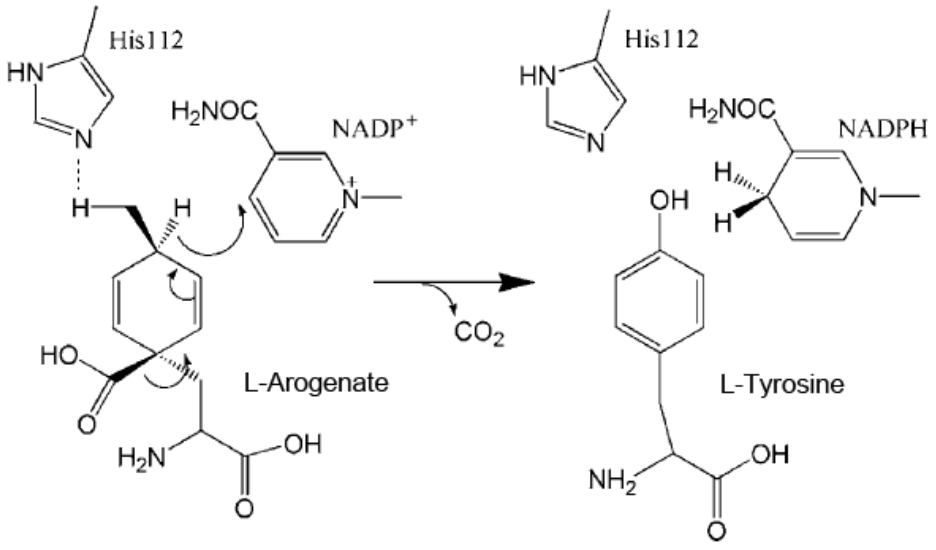


Figure 1.13: AD reaction mechanism. The concerted reaction mechanism as proposed by Hermes and coworkers [(85)]. His112 (*Synechocystis* sp. PCC 6803 AD numbering) is proposed to be a key catalytic H-bond acceptor in the dehydrogenase reaction. Adapted from Legrand and coworkers[(41)].

1.10 TyrA enzymes from hyperthermophilic archaea *Nanoarchaeum equitans* and *Ignicoccus hospitalis*

N. equitans is the smallest of the hyperthermophilic archaea characterized to date and possesses a highly reduced genome. The organism thrives through a unique obligatory symbiotic relationship with the larger archaeon *I. hospitalis* and grows optimally at 90°C [(119, 120)]. The crenarchaeon *I. hospitalis* is a hydrogen-oxidizing, chemoautotroph that couples CO_2 fixation with sulfur respiration. Uniquely among archaea, *I. hospitalis* cells are surrounded by two membranes separated by a periplasmic space within which vesicular transport occurs [(121)]. The potential involvement of this transport in the relationship with *N. equitans* is unknown. The

complete genomic sequences of *N. equitans* and *I. hospitalis* were reported in 2002 and 2008, respectively, isolated from a co-culture of *N. equitans* and *I. hospitalis* or from a pure culture of *I. hospitalis*, which were initially derived from samples found in Icelandic thermal vents[(53, 121)]. Genome analysis revealed that *N. equitans* has an A-T rich genome (69%) and lacks most of the genes for the biosynthesis of amino acids, nucleotides, lipids and cofactors as well as the pathways for carbon assimilation such as glycolysis, gluconeogenesis and the pentose phosphate pathway[(119)]. Accordingly, *N. equitans* must transport most cellular metabolites from its host organism *I. hospitalis*. Interestingly, *N. equitans* retains a trifunctional PD-CM-PDT gene (NEQ192) required for aromatic amino acid biosynthesis [(119)]. While most of the known TyrA enzymes are either monofunctional or a fusion of CM-PD domains, the *N. equitans* TyrA is predicted to possess all three enzymatic activities on one polypeptide chain. The unusual three domain assembly is evident in only two other organisms: the hyperthermophilic archaeon *Archaeoglobus fulgidus* which is reported to possess an NAD^+-specific, tyrosine-sensitive trifunctional PD [(55)] and the brown algae *Ectocarpus siliculosus* whose genome is predicted to encode a CM-PDT-PD fusion protein [(54)]. Genome analysis of *I. hospitalis* has identified a *tyrA* gene (gene locus Igni_0892) that is predicted to encode a 348- residue bifunctional CM-PD. PDT activity, in contrast, is believed to be encoded by a separate gene.

Amino acid sequence alignment of the CM and PD domains of *I. hospitalis and N. equitans* TyrA proteins has revealed only 36% and 24%, sequence identity, respectively. Additionally, the archaeal PD domains have only ~25% amino acid sequence identity with their homologues from *E. coli* and *A. aeolicus*. Thus, functional studies on the archaeal TyrA proteins would provide additional insights in the catalytic mechanism and mode of regulation of PD proteins. For example, while most PD proteins use NAD^+ as a cofactor in the dehydrogenase reaction[(29)], amino acid sequence analysis of a number of TyrA proteins predicts that both

symbiotic archaea possess *tyrA* genes that may give rise to $NADP^+$-specific PD activities (Fig 1.8). Moreover, some residues that are important in the coordination of L-Tyr in PD proteins from *E. coli* and *H. influenzae*, do not appear to be conserved in *I. hospitalis and N. equitans* TyrA. Thus, PDs from both symbiotic archaea may be unregulated by L-Tyr; this is unusual since most PD proteins are strictly feedback inhibited by the end product. The enzymes from both archaeal symbionts may represent a distinct class of TyrA proteins.

To date, all efforts to obtain a three-dimensional structure of multifunctional TyrA proteins have not been successful. However, several monofunctional CM and PD proteins have been crystallized and their structures have been solved [64, 73, 78, 89, 91, 92, 122, 123]. The physical properties of hyperthermophilic enzymes, such as a highly charged surface[124-126] and increased packing density[126, 127], are known to render them better candidates for crystallization studies. Accordingly, multifunctional TyrA proteins from the hyperthermophilic archaeal symbionts could be interesting candidates for crystallography studies.

The rare PD/CM/PDT fusion of the trifunctional protein, the apparent unusual cofactor specificity and possible L-Tyr deregulation of the hyperthermophilic TyrAs from both symbionts make these proteins good model systems for comparative structural and functional studies.

1.11 Scope and organization of thesis

Chapter two describes the recombinant expression, purification and characterization of *I. hospitalis* CM-PD and *N. equitans* PD-CM-PDT. *I. hospitalis* and *N. equitans tyrA* genes were cloned into pET-15b, expressed in *E. coli* and chromatographed on Ni-NTA resin. While the strategy was successful for obtaining satisfactory yields of *I. hospitalis* TyrA, the expression of *N. equitans* TyrA was poor and that which was expressed was mainly insoluble and copurified with

chaperone proteins that assist in proper protein folding, and proteins with affinity to nickel. We attempted to increase the yield of soluble recombinant *N. equitans* TyrA by exploring other *E. coli* expression systems, and additional chromatography steps during its purification. The *tyrA* gene of *N. equitans* was synthesized and the sequence was modified to improve the recombinant expression of the trifunctional enzyme. The levels of expression and purification of both archaeal TyrA proteins were analyzed by SDS-PAGE, 2D-PAGE and Western Blot analysis as well as by intact and tandem MS analysis. The ESI-MS detected the monomeric masses of the archaeal TyrA proteins while their native oligomeric state was evaluated using analytical size exclusion chromatography (SEC-FPLC) and/or analytical ultracentrifugation.

In chapter three, the effect of temperature on the activity and structure of *N. equitans* and *I. hospitalis* TyrA proteins was assessed using far-UV circular dichroism spectrometry and kinetic assays. We confirmed the presence of enzymatic activities predicted for both TyrA proteins, and assessed the effect of pH and salt on enzyme function. The kinetic parameters for the reactions were determined as well as regulation by the end products L-Phe and L-Tyr. The stability of *I. hospitalis* was further explored. Using limited proteolysis, intact and tandem MS analysis we identified a 30 kDa stable domain which we attributed to a degradation product comprising the C-terminal portion of the protein.

In chapter four, the characterization of selected *I. hospitalis* CM-PD variants was described as efforts to probe domain structure and specificity towards the binding of substrates and the end product L-Tyr. The stable domain identified in Chapters 2 and 3 that copurified with the wild-type enzyme was expressed in *E. coli* as a Δ80CM-PD variant (deletion of the first 80 amino acids from the protein). Met81Leu was also expressed to assess whether this 30 kDa fragment might be a result of an alternative start site leading to a shorter form of the enzyme. This variant instead provided information regarding spatial relationship between the CM and PD

domains. Variants Δ80CMPD and Met81Leu were purified via Ni-NTA chromatography and their kinetic and biophysical properties were determined.

Amino acid sequence alignment of PD proteins with the PD domain of *I. hospitalis* predicts $NADP^+$- dependent dehydrogenase activity for *I. hospitalis* TyrA (Fig 1.8 and Jensen *et al* [29]). To examine cofactor specificity, variants Gly126Asp, Arg127Leu/Ala and Asn128Ala/Asp were characterized. The amino acid sequence alignment (Fig 1.8) also revealed a conserved residue Arg308 which corresponds to Arg294 of *E. coli* CM-PD. The importance of Arg308 in binding prephenate was explored by characterizing Arg308/Lys/Ala/Gln variants. Additionally, it has been shown that Tyr285' and Tyr303 coordinate the amine of L-Tyr and are important for end product binding in *E.coli* CM-PD[94]. Accordingly, amino acid replacements Phe299Tyr and Ile316Tyr were introduced in *I. hospitalis* CM-PD as efforts to regain sensitivity to L-Tyr. All variants discussed in this chapter were expressed in *E. coli,* purified via Ni-NTA chromatography and their kinetic properties were determined.

Chapter 2

Production and Molecular Weight Determination of TyrA Proteins from *Nanoarchaeum equitans* and *Ignicoccus hospitalis*

2.0 Introduction

In this chapter we report the first heterologous expression and purification of putative TyrA proteins from the hyperthermophilic archaea *N. equitans and I. hospitalis. N. equitans* is the smallest hyperthermophilic archaeon characterized to date and thrives through a unique symbiotic relationship with the larger archaeon *I. hospitalis* from which it obtains most of the biomolecules (such as lipids, nucleotides and amino acids) required for life. The complete genomic sequences of *N. equitans* and *I. hospitalis* were reported in 2002 and 2008, respectively, isolated from a co-culture of *N. equitans* and *I. hospitalis* or from a pure culture of *I. hospitalis*, which were initially derived from samples found in Icelandic hydrothermal vents [53, 121]. The genome of *N. equitans* is compact and A-T rich and it is predicted to encode an unusual trifunctional TyrA protein (NeTyrA) with PD-CM-PDT assembly where all three enzymatic activities are found on one polypeptide chain[119]. In contrast, the *tyrA* gene from its host *I. hospitalis* (gene locus Igni 0892) is predicted to encode a 348-residue bifunctional CM-PD (IhTyrA).

Bioinformatic analysis revealed that the N-terminal part of the NeTyrA protein (residues 1-236) likely comprises the PD domain and is followed by a CM domain (residues 237-327) that is followed by the PDT domain (Fig 2.1). The PDT portion contains a putative ACT (aspartokinase-chorismate-mutase-TyrA) regulatory domain that is involved in allosteric regulation by L-Phe[29, 56]. Sequence analysis of IhTyrA revealed that the N-terminal part of the protein (residues 1-89) likely comprises the CM domain, which is followed by the PD domain (residues 90-348).

In this chapter we present the cloning, expression and the partial purification strategy for both archaeal TyrA proteins. The *tyrA* genes from *N. equitans* and *I. hospitalis* were cloned into a pET-15b vector, which permits the addition of a removable N-terminal hexa-His tag, expressed in *E. coli* and chromatographed on Ni-NTA resin. Their purification was assessed by

SDS-PAGE and mass spectrometry and by PD activity assays. While the affinity purification of IhTyrA was effective, the yield of NeTyrA was low and its expression and purification were accompanied by contaminating *E. coli* proteins. The gene for NeTyrA was synthesized, which resulted in an improvement in the yield of active enzyme. We confirmed the presence of all catalytic domains (CM, PD, PDT) associated with each protein by activity measurements at an elevated temperature. In this chapter mass spectrometry was extensively used to characterize the purification and properties of IhTyrA and NeTyrA enzymes, and to identify a stable protein fragment of the bifunctional IhTyrA protein generated during its expression and/or purification.

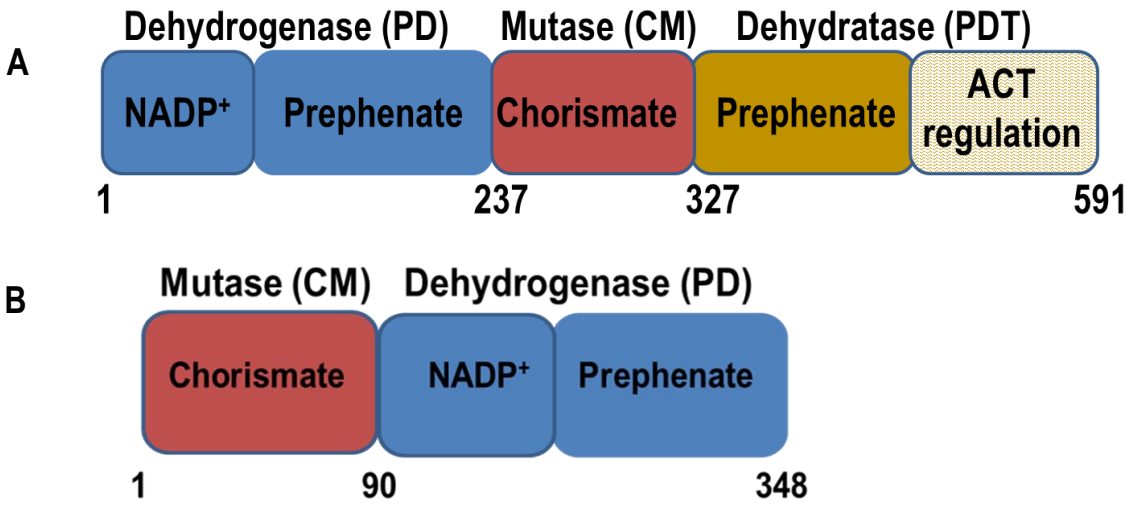


Figure 2.1: Predicted domain organization of (A) NeTyrA and (B) IhTyrA.

2.1 Experimental procedures

2.1.1 Materials

Prephenate (barium or sodium salt) and chorismate (free acid), were prepared as previously described [(128, 129)], while NAD^+ and $NADP^+$ (free acid) were obtained from Roche. Concentrations of stock substrate solutions were determined using published extinction coefficients [(130)] and/or enzymatic end-point analysis. All acids and organic solvents for mass

spectrometry were HPLC grade. Trifluoroacetic acid (TFA) was obtained from Sigma. Trypsin powder, for in-solution tryptic digestion (sequencing grade modified) and cOmplete®, EDTA-free protease inhibitor cocktail tablets were purchased from Roche. Ampicillin (sodium salt), kanamycin sulphate and chloramphenicol, Isopropyl β-D-1-thiogalactopyranoside (IPTG) and phenyl-methyl-sulfonyl fluoride (PMSF) were obtained from BioShop and stock solutions were prepared and stored as outlined previously[(131)]. Benzonase nuclease was purchased from Novagen. Thrombin was obtained from Sigma. Ni-NTA Superflow® chromatography resin was supplied by Qiagen. Dialysis membrane (12-14 kDa cut-off) was from Spectrapor and ultrafiltration units (30 or 10 kDa cutoff) were obtained from Amicon and were prepared according to manufacturers' instructions. Restriction enzymes, recombinant *PfuTurbo®* DNA polymerase (2.5 U/µL) and the deoxy-NTP (dNTP) mixture (5 mM of each dNTP, stored at -20°C in small aliquots) were purchased from MBI Fermentas. *Phusion®* High-Fidelity DNA Polymerase (2.0 U/µL) was purchased from New England Biolabs Inc. Oligonucleotides were obtained from Integrated DNA Technologies at standard purity. Rabbit anti-His antibodies were from Santa Cruz Biotechnology Inc while alkaline phosphatase- conjugated secondary antibodies and the color development substrates 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) and nitro blue tetrazolium (NBT) were from Sigma. All other chemical reagents were obtained commercially and were of the highest quality available.

2.1.2 Strains and plasmids

The *E. coli* strain XL10-Gold ultracompetent (Stratagene) Tet^rΔ(*mcrA*)*183 Δ*(*marCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F'*proAB lacI*^{q}*ZΔM15* Tn10 (Tet^r) Amy Cam^r] was used for plasmid production while BL21(DE3) (Stratagene) [F^- dcm^+ Hte ompT hsdS(r_B^- m_B^-) gal λ (DE3) endA Tet^r] and Rosetta™2(DE3) (Novagen) [F^- ompT $hsdS_B$(r_B^- m_B^-)

gal dcm (DE3) pRARE2 (Cam^R)] were used for protein expression. The cells were either purchased commercially ready-to-use or when required, laboratory stocks of these cells were made chemically competent using calcium chloride [131]. The helper plasmid, pMgK, was kindly donated by Dr. D. Christendat from the University of Toronto, Canada while *I. hospitalis* and *N. equitans* genomic DNA mixture (KIN4I/M) was generously provided by Drs. Robert Huber and Karl O. Stetter from the University of Regensburg, Germany. The expression vector pET-15b was kindly provided by Dr. P. Pawelek. Recombinant *E. coli* CM-PD was expressed and purified as described elsewhere[58] .

2.1.3 Cloning of *tyrA* genes into pET-15b

DNA manipulations were performed according to standard methods [131] or as suggested by manufacturers' protocols. *N. equitans tyrA* was amplified by PCR from a KIN4 genomic DNA mixture of *I. hospitalis* and *N. equitans* using the forward (P1) and reverse (P2) primers

P1: 5'AATTCATATGATGGTTATGATATTAATTATT3' (T_m = 47°C) and

P2: 5'GCGCGGATCCTTAATCTAGAATCTTAGGGAA3' (T_m = 52°C) while *I. hospitalis tyrA* was amplified from the KIN4 genomic DNA mixture using primers:

P1: 5'AATTCATATGAGCGAGAACCCGCTCGAGT3'(T_m = 71°C) and

P2: 5'TTAAGGATCCTTAACGGCCTTTCCTCCACCT3'(T_m = 67°C). PCR conditions used for the *tyrA* gene amplification are listed in Table 2.1.

Number of Cycles	Temperature (°C)		Time
1	Denaturation	95	2 min
35	Denaturation	95	30 sec
	Annealing	42 / 53	40 sec
	Extension	72	4 min
1	Extension	72	10 min
	Cooling	4	overnight

Table 2.1: Parameters used for PCR-mediated amplification of *tyrA* genes. An annealing temperature of 42°C was used to amplify the *N. equitans tyrA* gene while 53°C was used to amplify the *I. hospitalis tyrA* gene.

Forward primer, P1, carried the *Nde*I cleavage site (underlined). Reverse primers are denoted by P2 and carried the *Bam*HI cleavage site (underlined). Oligonucleotides were resuspended in MilliQ water to a final concentration of 100 µM. Their melting temperatures (T_m values) were calculated using New England Biolabs web site http://tmcalculator.neb.com/#!/. Reaction mixes of 50 µL consisted of 5 ng genomic template, 2.5 µM of each primer, 2 µL of 5 mM dNTPs, ddH_20 and 5 x *Pfu* reaction buffer supplemented with $MgSO_4$. *Pfu* DNA polymerase (0.5 µL/1.25 U) was added immediately before the first cycle to start the reaction. Annealing temperatures were chosen 5°C below the lowest melting temperature calculated for the oligonucleotides to allow sufficient annealing. PCR products were resolved on 1% agarose gels and purified using QIAEX II Gel Extraction Kit (Qiagen). The clean PCR products and plasmid DNA (pET-15b empty vector) were individually digested with *Nde*I and *Bam*HI (double digestion) using supplier-provided protocols. Linearized pET-15b was dephosphorylated by reaction with Calf Intestine Alkaline Phosphatase using the manufacturer-supplied protocol. Digested and

dephosphorylated pET-15b and the digested PCR products were then resolved on 1% agarose gels and purified using QIAEX II Gel Extraction Kit (Qiagen). Each purified digest of insert (*tyrA* gene) and dephosphorylated pET-15b were mixed together in a 4:1 ratio of insert to vector and incubated overnight at 16°C with T4 DNA ligase. The ligase was heat-inactivated at 65 ^{o}C as per manufacturer's instructions and the cloned DNA (*tyrA*-containing plasmids) was then transformed into *E. coli* strain XL10-Gold competent cells. Briefly, this transformation involved incubating 50 µL of XL 10-Gold competent cells with 5 µL of cloned DNA on ice for 10 min. The cells were then incubated at 42°C for 45 s (heat-shock) and returned to the ice for another 10 min. Growth of the transformed cells was selected on Luria-Bertani (LB) agar plates containing 100 µg/mL ampicillin. Single colony transformants were propagated in 10 mL of LB/Amp and the recombinant plasmid DNA extracted using a GeneJET^{TM} Plasmid Miniprep Kit. The extracted recombinant plasmids were sequenced at Genome Québec Innovative Center at McGill University and the resulting sequences were verified using the BLAST tool in NCBI (http://www.ncbi.nlm.nih. gov/blast) to ensure that no unwanted mutations had been introduced during the PCR reaction.

2.1.4 The *tyrA* gene synthesis

N. equitans tyrA (NEQ192) gene was synthesized by GeneArt® Life Technologies using a codon bias to optimize protein expression in *E. coli*. Additionally, nucleotide changes were incorporated to disrupt a hydrophobic stretch of residues at the protein's N-terminus: amino acids at the second and third position in the primary sequence (valine and methionine) were replaced by four residues (isoleucine, serine, valine and lysine) found in *tyrA* from *A. fulgidus*. The modification of N-terminus is underlined: MGSSHHHHHHSSGLVPRGSHMISVKILIIGFGRLGQYFY…

The optimized *tyrA* gene was cloned into pET-15b and the sequence confirmed by DNA sequence analysis.

2.1.5 Expression of recombinant NeTyrA and IhTyrA

Cloned NeTyrA was overexpressed in *E. coli* Rosetta2(DE3) cells while the TyrA from the synthesized gene was overexpressed in *E. coli* strain BL21(DE3). Recombinant IhTyrA was expressed in *E. coli* BL21(DE3) cells harboring plasmid pMgK. The *E. coli* cells were transformed with plasmid containing the *tyrA* gene and cultured in LB agar plates supplemented with either 100 µg/mL ampicillin (when using BL21(DE3 cells) or 100 µg/mL ampicillin and 30 µg/mL chloramphenicol (when using Rosetta2(DE3) cells) or 100 µg/mL ampicillin and 50 µg/mL kanamycin (when using pMgK harboring BL21(DE3) cells). After overnight growth at 37°C, a colony from each plate was transferred to 50 mL of LB medium supplemented with appropriate antibiotics and grown at 37°C for 16 h with shaking. Bacterial cultures of ten mL were then diluted into 1 L of the same medium and were grown at 37°C to an OD_{600} of 0.6. Gene expression was then induced by the addition of 0.2 or 0.4 mM IPTG and cells were incubated overnight with shaking at 18°C and harvested by centrifugation.

2.1.6 Preparation of cell lysate

IhTyrA and NeTyrA proteins were chromatographed on Ni-NTA affinity resin although in the case of NeTyrA additional purification methods were evaluated. For large scale cultures frozen cell pellets were thawed and resuspended in ice cold purification buffer (50 mM Tris, 500 mM NaCl, 5% glycerol at pH 8, in the case of NeTyrA the buffer also contained 10 mM β-mercaptoethanol) supplemented with cOmplete® protease inhibitor cocktail (Roche, one tablet per 50 mL suspension), 0.5 mM phenylmethyl sulfonyl fluoride (PMSF) and benzonase

nuclease. The cells were resuspended by ten up-and-down strokes of a Dounce Homogenizer and cell walls disrupted by two passages through a French Press (Thermo Spectronic) at 18 000 psi. The membrane fraction was removed by centrifugation at 45 000 x g for 45 min at 4°C. For smaller scale expression and purification studies, cell walls were disrupted by sonication (Biologics Inc) using microtip at 10 bursts X 10 sec (power settings of 40%) with 1 min interval on ice between bursts. The cell lysate was then centrifuged in an Eppendorf Microfuge (highest speed). Protein levels in the crude extract, cell-free extract and pellet were evaluated by SDS-PAGE analysis.

2.1.7 Chromatography of NeTyrA and IhTyrA using Ni-NTA affinity resin

Cell-free extract was applied to Superflow Ni-NTA resin (Qiagen, column bed volume 10 mL) equilibrated with purification buffer containing 5 mM imidazole. Cleared lysate was cycled through the resin overnight at 4°C. The resin was extensively washed with purification buffer supplemented with 30 mM imidazole and bound protein was eluted with a 100-300 mM imidazole stepwise gradient. For the stepwise elution, the column was washed with purification buffer containing 30, 100, 150, 200, 250 mM and 300 mM imidazole (2 column volumes each) and 5 mL fractions were collected. Active NeTyrA protein eluted at 100-200 mM imidazole while IhTyrA eluted at 150-300 mM imidazole. For chromatography of cloned NeTyrA, the elution of bound protein was performed with buffer containing 400 mM imidazole and fractions of 1 mL were collected. Elution fractions were supplemented with 1 mM EDTA and in the case of NeTyrA with 1 mM dithiothreitol (DTT). All chromatography steps were performed at 4°C. Fractions were analyzed by SDS-PAGE (12% polyacrylamide) and PD activity measurements. Appropriate fractions, selected by SDS-PAGE and PD activity, were pooled and dialyzed overnight in storage buffer (50 mM Tris, 300 mM NaCl, 1 mM EDTA 5% glycerol, pH 8) at 4°C. In the case of NeTyrA,

the buffer was supplemented with 1 mM dithiothreitol (DTT). If required, proteins were concentrated (Amicon Ultra-15, 30 kDa cut-off), and all preparations were stored at -80°C in storage buffer supplemented with 20% glycerol (NeTyrA solutions also contained 10 mM DTT). To remove the hexa-His tag, thrombin (GE Healthcare) was added to purified TyrA at a ratio of protein:thrombin of 100:1 (w/w). Samples were dialyzed overnight at room temperature in storage buffer containing 2.5 mM $CaCl_2$ and then reapplied onto the Ni-NTA column. Unbound protein was collected, concentrated and stored as described above.

2.1.8 Evaluation of heat treatment step for purification of NeTyrA

A cell-free extract from a small culture of NeTyrA was prepared in purification buffer containing 1 M NaCl. Samples were heated at 80°C for 20 min, transferred to ice and centrifuged (30 000 x g, 10 min, 4°C) to remove precipitated matter. Proteins in the supernatant were resolved by SDS-PAGE on a 12% polyacrylamide gel.

2.1.9 Additional column chromatography of NeTyrA

Ni-NTA affinity purified NeTyrA (from 12 L culture) was further chromatographed in parallel on cation exchange, anion exchange and phenyl sepharose resin on a small scale (5 mg protein sample) following procedures used for *E. coli* CM-PD and/or resin manufactor-supplied protocols. Additionally, 10 mg of protein was also chromatographed on Red Agarose 120 $NAD(P)^+$ affinity resin (25 mL) according to the method of Legrand *et al* that was reported for the purification of *Synechocystis sp.* PCC 6803 AD [41]. Protein from another large culture (6 L) was chromatographed on Q- Sepharose anion exchange resin (50 mL, equilibrated with 50 mM Tris-HCl, 1 mM DTT at pH 7.5) using a 0-2 M NaCl gradient. After dialysis overnight, the protein was rechromatographed on Ni-NTA affinity resin as indicated in 2.1.7. None of these trials were

particularly effective and are not discussed in detail in this thesis. The exception was the protein obtained from anion exchange followed by Ni-NTA chromatography, which was analyzed as discussed in the text.

2.1.10 SDS-polyacrylamide gel electrophoresis and 2D gel analysis

Denaturing SDS-PAGE was performed with either a 12% or 15% polyacrylamide gel following the method of Laemmli[(132)]. Protein samples were diluted 5:1 (v/v) into 5 X SDS sample loading buffer (1.5 M Tris-HCl, 4% SDS, 20% glycerol, (v/v), 5% β-mercaptoethanol, 0.002% Bromophenol Blue, pH 6.8) and boiled with loading buffer for 10 min, centrifuged and cooled on ice 2 min prior to loading on a gel. The sample was electrophoresed at 120 to 200 V. Electrophoresis continued until the Bromophenol Blue tracking dye migrated off the resolving gel. Bio-Rad broad range molecular weight proteins standards were used to estimate the molecular weight of proteins in the samples. Protein was visualized by Coomassie Brilliant Blue R-250 staining or by using the Pierce™ silver stain kit.

2D gel electrophoresis is a method that resolves proteins according to their isoelectric point via isoelectric focusing (IEF) followed by separation according to mass via SDS-PAGE. 2D gel electrophoresis was performed according to BioRad protocol for 2D gel analysis (catalog number 163-2105). Briefly, protein samples that were concentrated previously by acetone precipitation as outlined by Crowell and coworkers [(133)] and stored at -20°C, were resuspended in sample/rehydration buffer (BioRad) and then a 125 µL aliquot was loaded into the Isoelectric focusing (IEF) tray. The 7 cm immobilized pH gradient (IPG) strips (pH 3-10) were placed face - down on the protein sample, covered and were actively rehydrated overnight (50 V). The following day IEF was performed with the rehydrated IPG strips using the program setup as shown: (1) 150 V, 30 min, (2) 300 V, 30 min, (3) 600 V, 30 min, (4) 5000 V, 2.5 h, (5) 8000 V,

35000 Vh, (150 V, 24 h). The next day the IPG strips were washed for 20 min with equilibration buffer 1 (6 M Urea, 2% SDS, 0.375 M Tris-HCl pH 8.8, 20% glycerol and 2 % DTT) followed by a wash with equilibration buffer 2 (6 M Urea, 2% SDS, 0.375 M Tris-HCl pH 8.8, 20% glycerol and 2% iodoacetamide) to ensure reduction and alkylation of cysteine residues. The washed IPG strips were then washed briefly with SDS running buffer and placed on the resolving SDS-polyacrylamide gel, overlaid with agarose and electrophoresed by SDS-PAGE as described above. The resulting gel was visualized by Coomassie Brilliant Blue R-250 staining.

2.1.11 Western blot analysis of His-tagged proteins

Expression of soluble His-tagged protein was confirmed by Western blotting with Anti-His antibodies (Santa Cruz Biotechnology) using an alkaline phosphatase-conjugated secondary antibody (Sigma). Briefly, protein samples were resolved by SDS-PAGE as described in 2.1.10 and wet transferred onto nitrocellulose membranes (prewashed with transfer buffer of 25 mM Tris, 195 mM glycine, 20% methanol, pH 8.3) at 10 V for 1 h. As a control, separate gels were also stained with Coomassie Blue to visualize protein bands. After blocking for 1 h with 1 % bovine serum albumin (BSA) in TBS (150 mM NaCl, 50 mM Tris, pH 7.5), membranes were incubated with a primary antibody (Anti-His Ab diluted 1:3000 dilution) for 1 h, washed 3 x 5 min with TTBS (20 mM Tris, 500 mM NaCl, Tween-20 pH 7.5) and incubated with alkaline phosphatase-conjugated secondary antibody (1:3000 dilution) for 45 min. For color development, the secondary antibody was removed, and membranes were washed 3 x 5 min with TTBS, 5 min with alkaline phosphatase (AP) buffer (100 mM Tris, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) and BCIP/NBT color development solution was added. The BCIP/NBT color development solution was prepared by adding 60 µl of 50 mg/mL BCIP in 100% dimethylformamide (DMF) and 120 µl

of 50 mg/mL NBT in 70% DMF to 20 mL of AP buffer. When the desired color development was achieved, the reaction was stopped by several changes of distilled water.

2.1.12 Determination of protein concentration

Protein concentration was quantified using the Bradford method with the Bio-Rad protein assay kit[134] (Bio-Rad Laboratory) with bovine serum albumin (BSA, Sigma) as a standard. BSA was prepared in 10 mM Tris-HCl, pH 7.4 and filtered using 0.2 µm filters. Its concentration was determined by OD_{280} readings using an extinction coefficient of 0.667 mL/mg/cm.

2.1.13 Determination of enzyme activity

Prephenate dehydrogenase activity was recorded by continuously monitoring the formation of NADH (or NADPH) in the presence of NAD^+ (or $NADP^+$) and prephenate at 340 nm, while the conversion of chorismate to prephenate catalyzed by chorismate mutase was determined by following the decrease of chorismate at 274 nm as described in Turnbull *et al*[26]. Both CM and PD reactions were monitored continuously using a Varian Cary 50 spectrophotometer equipped with a thermostatically regulated cuvette holder. CM and PD activities were routinely assayed at 80°C in reaction buffer containing 50 mM HEPES, 200 mM NaCl (pH 8 when titrated at ambient temperature). The PD reaction was measured in 1 mL quartz cuvettes (path length 1 cm) while the CM reaction was measured in a 0.5 mL quartz cuvettes (path length 0.5 cm). Since both chorismate and prephenate are heat labile [130], the degree of thermal decomposition over the course of the assay was determined and if required, the kinetic values were corrected. Reaction buffer was preincubated for 2 min at the appropriate temperature, and then substrates (at room temperature) were added. After 15 s, the reaction was initiated by the addition of enzyme. Initial velocities were calculated from the linear portion of the progress curves from which non-enzymatic rates were subtracted. These calculations were

performed using the Cary WinUV kinetics software. For calculations of initial velocities an extinction coefficient of 6400 $M^{-1}cm^{-1}$ was used for the PD reaction (which takes into account the contribution of 4-hydroxyphenylpyruvate at 340 nm) and a value of 2630 $M^{-1}cm^{-1}$ was used for the CM reaction[87]. A unit of enzyme was defined as the amount of enzyme required to produce 1 µmol of product per min at the specified temperature.

Prephenate dehydratase activity was determined with a stopped time spectrophotometric assay following the appearance of phenylpyruvate at 320 nm [104]. Assays were conducted at 80°C as described above. The reaction was quenched by the addition of 1 M NaOH. Samples were then chilled and absorbance readings were recorded at room temperature. Reaction times were adjusted such that less than 20% of substrate was converted to product and non-enzymatic rates were subtracted from these values. Units of activity and specific activity were calculated as outlined below:

Calculation of mutase activity in units (µmol/min/mL):

$$\text{Units (µmol/min/mL)} = \frac{\Delta OD_{273}/\text{min}}{0.5\text{ cm} \times 2630\ M^{-1}\ cm^{-1}} \times \frac{10^6\ \mu\text{mol}}{\text{Mol}} \times \frac{1\text{L}}{10^3\ \text{mL}} \times \text{dilution factor}$$

Calculation of dehydrogenase activity in units (µmol/min/mL):

$$\text{Units (µmol/min/mL)} = \frac{\Delta OD_{340}/\text{min}}{1\text{ cm} \times 6400\ M^{-1}\ cm^{-1}} \times \frac{10^6\ \mu\text{mol}}{\text{Mol}} \times \frac{1\text{L}}{10^3\ \text{mL}} \times \text{dilution factor}$$

Calculation of dehydratase activity in units (µmol/min/mL):

$$\text{Units (µmol/min/mL)} = \frac{\Delta OD_{320}/\text{min}}{1\text{ cm} \times 17600\ M^{-1}\ cm^{-1}} \times \frac{10^6\ \mu\text{mol}}{\text{Mol}} \times \frac{1\text{L}}{10^3\ \text{mL}} \times \text{dilution factor}$$

$$\text{Specific activity (µmol/min/mg)} = \frac{\text{Units}}{\text{mg/mL protein}}$$

2.1.14 Mass spectrometry

2.1.14.1 Determination of subunit molecular weights by ESI-MS

Subunit molecular weight was determined by electrospray ionizing mass spectrometry (ESI-MS). Analysis was carried out on a Waters Micromass Q-ToF-2 mass spectrometer operating in positive ion mode and interfaced with an Agilent HP 1200 HPLC. Ten µL of protein sample in 0.1% formic acid were injected onto a GRACE Vydac 214 MS C4 column (2.1 x100 mm) that was pre-equilibrated with 10 % acetonitrile and 0.1% formic acid. Protein was eluted from the column into the mass spectrometer using a gradient of 10-90% acetonitrile at a rate of 0.25 mL/min and the elution profile was monitored by UV detection at 280 nm. Mass spectrometer parameters were as follows: source block temperature 80°C; desolvation temperature 350°C; capillary voltage, 3.5 kV; cone voltage 80 V; RF lens 50 V; ToF, 9.1 kV. Data analysis and deconvolution were performed using MassLynx 4.0 software (Waters Micromass). Calibration of the instrument was performed with bovine serum albumin. Mass shifts of ± 3 mass units for TyrA are within the expected experimental error.

2.1.14.2 Tandem mass spectrometry analysis of tryptic-generated peptides from NeTyrA and IhTyrA

<u>In-gel tryptic digestion</u>

Tryptic peptides were obtained by resolving TyrA preparations by SDS-PAGE, excising appropriate regions from the gel and then performing trypsin proteolysis on each sample using the procedure outlined in Shevchenko *et al* [(135)]. Briefly, protein sample was resolved by SDS-PAGE (12% polyacrylamide) and visualized by Coomassie Blue or silver staining as indicated in section 2.1.10. The bands representing protein product were excised from the gel, cut to smaller pieces (<1 mm^2) and each piece was placed in an individual Eppendorf tube. The gel pieces were

then destained, reduced with dithiothretiol, alkylated with iodoacetomide, and digested with trypsin (ratio (w/w) of trypsin:protein of 1:20) overnight at 37 °C. The resulting peptides were extracted from the gel using aqueous and organic extraction steps (MilliQ water and 50% acetonitrile), dried completely in a SpeedVac and resuspended in 15-50 µL of solubilization solution containing 2% acetonitrile (ACN) and 1% of formic acid (FA) analyzed by LC-MS/MS.

Liquid chromatography-tandem MS

Liquid chromatography-tandem MS (LC-MS/MS) analysis was performed on tryptic peptides of TyrA samples using a Thermo EASY nLC II LC system coupled to a Thermo LTQ Orbitrap Velos mass spectrometer equipped with a nanospray ion source. Samples with peptides obtained from in-gel tryptic digestion as outlined above were processed by LC-MS/MS as described in Kathiresan *et al* [(136)]. Briefly, two µL of each sample was injected onto a 10 cm × 75 µm column packed in-house with Phenomenex Jupiter C18 stationary phase (3 µm particle diameter and 100 Å pore size) that was equilibrated with 5% ACN and 0.1% FA. Peptides were eluted into the ESI source using a 38-min 5-90% ACN gradient at a flow rate of 400 nL/min with mobile phase A (5% ACN and 0.1% FA) and B (97% ACN and 0.1% FA). A full MS spectrum (m/z 400-1400) was acquired in the Orbitrap at a resolution of 60000, and the ten most abundant multiple charged ions were selected for MS/MS sequencing in the linear trap with the option of dynamic exclusion. Peptide fragmentation was performed using a collision-induced dissociation at a normalized collision energy of 35% with an activation time of 10 ms.

MS/MS fragments were analyzed using Thermo Proteome Discoverer software (v1.3) with the SEQUEST search engine with mass filters for oxidation of methionine (+15.995 Da) and for alkylation of cysteine (+57.021 Da). The database was searched against a FASTA file containing the amino acid sequence of either His-tagged NeTyrA or IhTyrA (FASTA files of the untagged putative

proteins were from the NCBI website http://www.ncbi.nlm.nih.gov) and the complete proteome of *E. coli* strain K12 (Uniprot, UP000000625). The enzyme selected for the database search was trypsin with a maximum of three missed cleavage sites. Mass tolerances of the precursor ion and fragment ion were set at 10 ppm and 0.7 Da, respectively. Only peptides with high confidence (false discovery rate less than 1%) and proteins with at least three identified unique peptides were reported.

2.1.15 Analytical size exclusion chromatography

The apparent molecular weight of native TyrA proteins was determined at ambient temperature by analytical size exclusion chromatography (SEC) using a BioRad DuoFlow FPLC system fitted with a Superdex GE-200 column (HR 10/30, Pharmacia). Chromatography was carried out with mobile phases containing 50 mM potassium phosphate, 150 mM NaCl (pH 7.5) in the presence and absence of ligands at a flow rate of 0.75 mL/min and injection volume of 500 µL. Protein elution was monitored at 280 nm and fractions (0.5 mL) were assayed for PD activity. BioRad gel filtration protein standards included equine myoglobin (17 kDa), chicken ovalbumin (44 kDa), bovine γ-globulin (158 kDa), and thyroglobulin (670 kDa). Void volume and total bed volume were evaluated with Blue Dextran (2000 kDa) and vitamin B_{12} (1.35 kDa), respectively.

2.1.16 Analytical ultracentrifugation

Native molecular weight and shape of IhTyrA was also determined by sedimentation velocity experiments performed in a Beckman XL-I analytical ultracentrifuge and an An60Ti rotor using absorbance detection (280 nm). Purified IhTyrA protein samples were dialyzed at 4°C into a buffer containing 50 mM Tris, 200 mM NaCl and 1 mM EDTA (pH 8), diluted in the same buffer to give a final OD_{280} of 1 or 0.1 and loaded into 1.2 cm path-length doublesector charcoal-filled epon

centerpieces. Samples were centrifuged at 35 000 rpm and 20°C for 14 h. Values for the sedimentation coefficient (s) and an average molar mass were calculated from the velocity and shape of the sedimenting boundary using the program Sedfit (http://www.analyticalultracentrifugation.com/default.htm). The program Sednterp[(137)] was used to calculate partial specific volume of the protein, and buffer density and viscosity and to convert to $S_{20,w}$.

2.2 Results

2.2.1 Cloning of *tyrA* gene in pET-15b and expression strategy

N. equitans and *I. hospitalis tyrA* genes were successfully amplified individually from a genomic KIN4 DNA mixture of *I. hospitalis* and *N. equitans* by polymerase chain reaction (PCR) using the conditions outlined in section 2.1.4. Cloning of these genes into pET-15b allowed heterologous expression of TyrA proteins in *E. coli* cells. Proteins derived from this construct contained an N-terminal hexa-His tag (to aid with protein purification by Ni-NTA affinity chromatography), and a thrombin recognition sequence (to allow removal of the His tag by cleavage with the protease, after purification) (see Fig 2.2). To boost expression levels of NeTyrA in *E. coli,* the Rosetta™2 strain was used. This *E. coli* strain harbors plasmid pRARE2 that encodes tRNAs recognizing for seven codons: ($AGA^{(Arg)}$, $AGG^{(Arg)}$, $AUA^{(Ile)}$, $CUA^{(lue)}$, $GGA^{(Gly)}$, $CCC^{(Pro)}$ and $CGG^{(Arg)}$) that are frequently used by archaea and eukaryotes but are rare in *E. coli*. This strategy was successfully used to express proteins such as glutamate/aspartate-prephenate aminotransfarase from *Arabidopsis thaliana*[(138)] and archaeal ribonucleoproteins[(139)]. To boost the expression levels of IhTyrA, BL21(DE3) cells were co-transformed with the helper plasmid pMgK which encodes tRNAs recognizing rare codons in *E. coli*: $AUA^{(Ile)}$, $AGA^{(Arg)}$ and $AGG^{(Arg)}$. This

strategy was successfully used to express the monofunctional PD from the hyperthermophilic bacterium *A. aeolicus*[89].

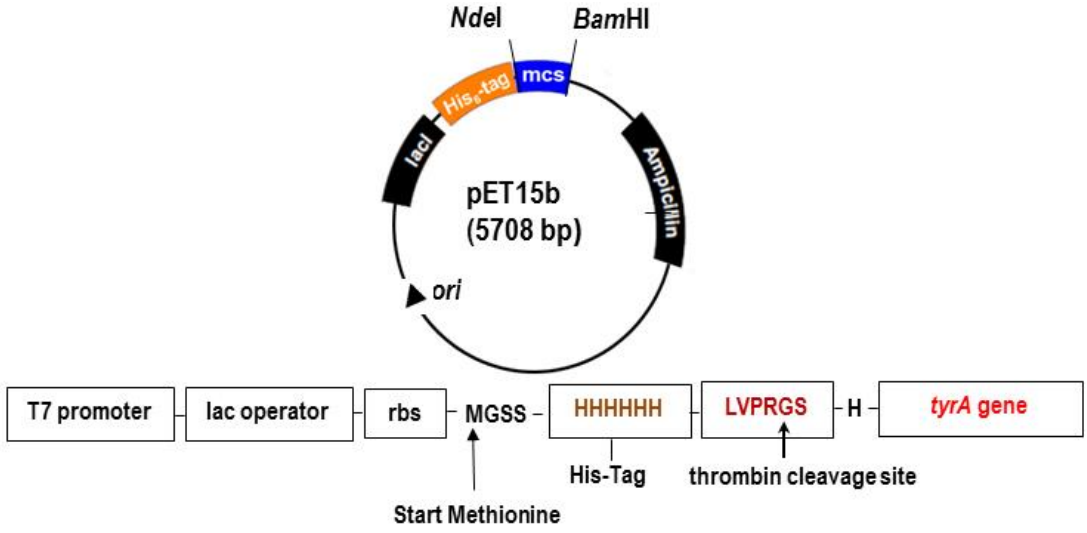


Figure 2.2: Novagen pET-15b expression vector. The *tyrA* gene encoding either *N. equitans* PD-CM-PDT or *I. hospitalis* CM-PD was cloned in-frame into the multi-cloning site between the *Nde*I and *Bam*HI restriction sites. This cloning allows inducible expression of the recombinant proteins with a thrombin-cleavable N-terminal hexa-histidine tag which facilitates purification via immobilized metal affinity. The primary sequence of hexa-His and thrombin recognition sites at the N-terminus of PD is shown. The two arrows identify the initiating Met of the open reading frame and site of thrombin cleavage. pET-15b diagram was adapted from (http://www.aidsreagent.org/pdfs/pet15b.pdf).

The expression level of TyrA proteins was first evaluated from a 10 mL LB medium culture supplemented with appropriate antibiotics. Protein expression was induced by the addition of IPTG, allowing accumulation for either 3.5 h at 37°C or overnight at 18°C. Cells were disrupted by sonication in the presence of protease inhibitors and the protein samples were analyzed by SDS-PAGE. As illustrated in Figure 2.3, a prominent band corresponding to ~71 kDa was obtained, which corresponds to the predicted mass for full-length NeTyrA. However, only a

small fraction of the overexpressed protein of interest appeared to be soluble. Overnight induction at 18°C revealed slightly better expression of a soluble protein than the induction at 37°C. Similarly to NeTyrA, the overexpression of IhTyrA was achieved as deduced by a prominent band at ~ 42 kDa (the predicted mass of full-length IhTyrA), although most of the overexpressed protein was insoluble (data not shown, see large scale expression analysis in 2.2.2). Induction with 0.2 mM IPTG for a longer time (72 h) at a lower temperature (18°C) did not appear to improve the expression. Thus, large scale protein expressions were conducted at 18°C using 0.4 mM IPTG, which was followed by Ni-NTA chromatography of the soluble protein.

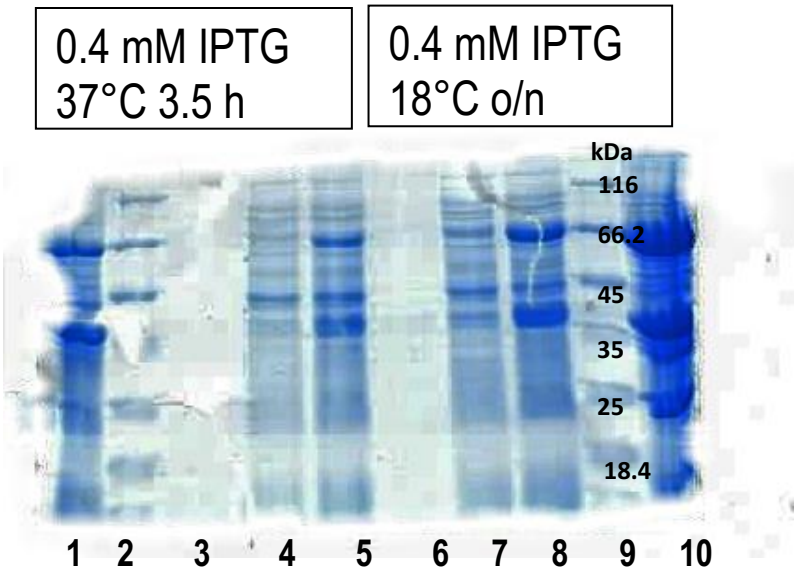


Figure 2.3: SDS-PAGE analysis of NeTyrA expression. Lane 1: Pellet of insoluble cellular debris solubilized in 4% SDS, Lane 2: Protein Molecular Weight Marker Lane 3: Heat step at 80°C, Lane 4: Cell-free extract, Lane 5: Cell-lysate, Lane 6: Heat step at 80°C, Lane 7: Cell-free extract, Lane 8: Cell-lysate, Lane 9: Molecular Weight Marker, Lane 10: Pellet of insoluble cellular debris solubilized in 4% SDS.

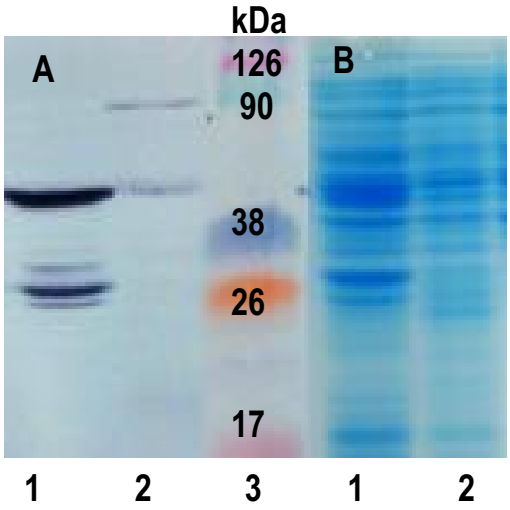


Figure 2.4: Western blot analysis of NeTyrA and IhTyrA expression. SDS-PAGE-resolved protein samples were: A. transferred onto membrane and stained with anti-His antibodies or B. stained with Coomassie Blue. Lane 1: Cell-free extract with IhTyrA, Lane 2: Cell-free extract with NeTyrA, Lane 3: Kaleidoscope Protein Molecular Weight Marker. Western blot analysis was performed as described in section 2.1.11. NeTyrA sample was from expression for 72 h at 18°C with 0.2 mM IPTG.

The expression of soluble His-tagged TyrA protein was also assessed by Western blotting with Anti-His antibodies. The immunoblot staining of cell-free extract from *E. coli* cells expressing the archaeal TyrA proteins revealed that the expression levels of soluble NeTyrA were much smaller than those of IhTyrA (Fig 2.4).

In parallel with these experiments other expression systems were evaluated in our efforts to improve the yield of soluble NeTyrA in *E. coli*. This included the use of: (1) pTRC99a, which is under the trc promoter and yields a non tagged protein used by Aitken *et al*[140]. We subcloned *tyrA* into pTRC99a, however no expression was observed; (2) pGEX , which is under the tac promoter and adds a removable glutathione-S transferase (GST) tag to facilitate purification by elution with glutathione as used by Leibovitch *et al* [141] *TyrA* was subcloned into the pGEX vector from pTRC99a, however no expression was observed.

2.2.2 Heat treatment to assist protein purification

Heat treatment is an effective method often used to purify thermostable proteins from *E. coli* cell-free extract[52, 55]. A cell-free extract of an *E. coli* harbouring plasmid containing *N. equitans tyrA* was heated in the presence of 1 M NaCl at 80°C as outlined in section 2.1.8, centrifuged and analysed by SDS-PAGE. As shown in Fig 2.3, lanes 3 and 6, many contaminating *E. coli* proteins were removed by this heat step including proteins which migrated at a molecular weight of ~ 71 kDa on the denaturing gel. Under these conditions we assumed that heating the sample may have removed NeTyrA along with the *E. coli* proteins.

2.2.3 Protein purification by Ni-NTA affinity chromatography

In order to obtain sufficient amount of soluble protein, large scale expression was conducted using a 2 L bacterial culture for IhTyrA and 12 L for NeTyrA expression. The His-

tagged proteins were chromatographed on Ni-NTA affinity resin as outlined in section 2.1.7 and summaries of the purification of TyrA from *N. equitans and I. hospitalis* are presented in Tables 2.2 and 2.3, respectively. SDS-PAGE analyses of fractions collected at various stages of the purification are shown in Figures 2.5 and 2.6, respectively.

fraction	concentration (mg/mL)	volume (mL)	total protein (mg)	total PD activity (U)	PD S.A. (U/mg)	yield PD (%)	purification (fold)
crude extract	6.9	275	1897	1243	0.65		
cell free extract	8.3	250	2075	353	0.16	100	
flow through	5.9	250	1475	90	0.06	25.5	0.4
pooled	2.5	13	32.5	40	1.20	11.3	7.4
dialyzed & stored	2.0	17	32.5	40	1.20	11.3	7.4

Table 2.2: Purification table of NeTyrA. Protein was expressed and purified from a 12 L cell culture. Protein production induced with 0.4 mM IPTG followed by 16 h growth at 18°C. Prephenate dehydrogenase activity was determined at 80°C in the presence of 0.5 mM prephenate and 2 mM NAD^+ (non-enzymatic rate subtracted). CM and PDT activities were also detected with 0.5 mM substrate and were 0.4 and 1.34 U/mg, respectively.

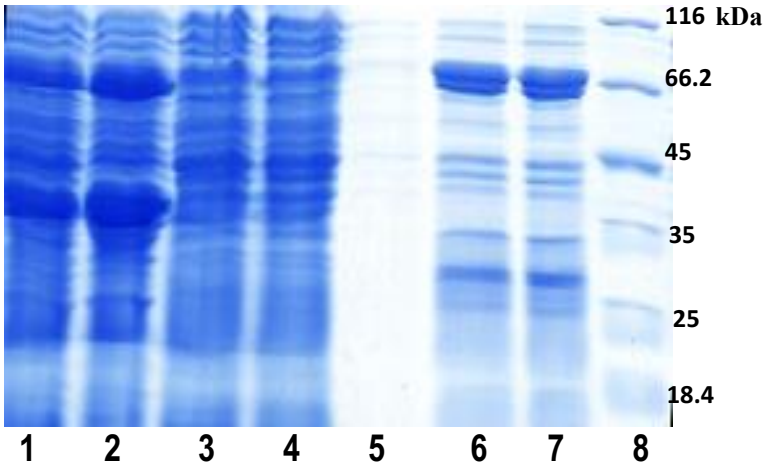


Figure 2.5: SDS-PAGE analysis of NeTyrA purification by Ni-NTA chromatography. Lane 1: Cell-lysate, Lane 2: Pellet of insoluble cellular debris solubilized in 4% SDS, Lane 3: Cell-free extract, Lane 4: Ni-NTA column flow through, Lane 5: 30 mM imidazole wash, Lane 6: Pooled protein, Lane 7: Dialyzed and stored protein, Lane 8: Molecular Weight Marker.

fraction	concentration (mg/ml)	volume (mL)	total protein (mg)	total PD activity (U)	PD S.A. (U/mg)	yield PD (%)	purification (fold)
crude extract	7.3	70	511	480	0.94		
cell free extract	5.1	70	356	436	1.22	100	1
flow through	2.5	85	210	132	0.63	30	0.5
pooled	0.6	52	30	202	6.71	46	5.5
concentrated & stored	3.2	6.5	21	114	5.4	26	4.4

Table 2.3: Purification table of IhTyrA. Protein was expressed and purified from a 2 L cell culture. Protein expression and prephenate dehydrogenase activity determination was performed as described in Table 2.2. CM activity was also detected and value of 9.1 U/mg was obtained with 0.5 mM chorismate.

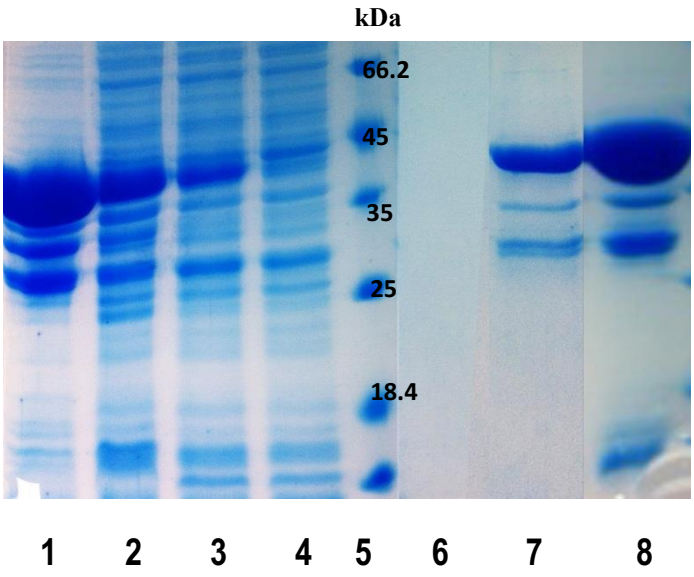


Figure 2.6: SDS-PAGE analysis of IhTyrA purification by Ni-NTA chromatography Lane 1: Pellet of insoluble cellular debris solubilized in 4% SDS, Lane 2: Cell-lysate, Lane 3: Cell-free extract, Lane 4: Ni-NTA column flow through, Lane 5: Protein Molecular Weight Marker, Lane 6: 30 mM imidazole wash, Lane 7: Pooled protein, Lane 8: Concentrated and stored protein.

SDS-PAGE analysis confirmed overexpression of protein corresponding to molecular weights of ~ 71 kDa and ~ 42 kDa, which is in agreement to subunit molecular weights of NeTyrA (Fig 2.5, lane 1) and IhTyrA (Fig 2.6, lane 2), respectively. Although much of the overexpressed protein was insoluble (Fig 2.5, lanes 2-3 and Fig 2.6, Lanes 1-3), nevertheless, soluble active TyrA was detected by assaying for PD specific activity (Tables 2.2 and 2.3). SDS-PAGE analysis of the IhTyrA protein showed a minor contaminant at ~ 30 kDa (Fig 2.6, lane 7) that was identified later as a degradation product comprising the C-terminal portion of the protein (see section 2.2.7). We attempted to eliminate this contaminant by washing the protein bound to Ni-NTA resin with buffer containing 1.5 M guanidine HCl. This was not effective however.

The preparation of NeTyrA contained many protein contaminants including those that appeared to migrate at a similar monomeric size to TyrA (doublet observed at ~ 71 kDa, (Fig 2.5 lanes 6 and 7) in the purified fraction. Interestingly, proteins eluting in the flow-through and washes of imidazole during the chromatography of both IhTyrA and NeTyrA possessed detectable specific activities despite their heterogeneity on SDS-PAGE, suggesting that these fractions contained different forms of PD that were catalytically active. The final yield of PD activity was 12% for NeTyrA was approximately 11% (Table 2.2) and 26% for IhTyrA. As an attempt to purify the NeTyrA further, the protein sample was treated with thrombin to cleave the hexa-His tag and then the thrombin-treated protein sample was applied the Ni-NTA resin. This strategy was not effective, however, and resulted in a very low yield of the trifunctional enzyme with reduced activity.

2.2.4 Additional purification strategies and evaluation by 2D gel electrophoresis

In an effort to purify NeTyrA from protein contaminants, we tested the utility of other conventional chromatography methods. The isoelectric point (pI) of NeTyrA predicted from its primary sequence at ~ 8.8 is quite high and distinct from many proteins including other TyrAs. Additionally, PD activity of the archaeal TyrA protein was shown to be cofactor dependent, a characteristic exploited during the purification of other TyrA proteins[(41)]. TyrA's overall hydrophobicity was also explored[(55)]. In our hands, however, High-S support cation exchange (BioRad), Phenylsepharose, or Reactive Red agarose $NAD(P)^+$ affinity resins were not particularly effective. A preparation of NeTyrA that appeared more homogeneous was obtained by anion exchange followed by Ni-NTA chromatography and eluted in a wash with high imidazole after the removal of the tag and reapplication onto the Ni-NTA column.

In order to assess the level of heterogeneity in the sample purified by anion exchange and Ni-NTA chromatography we used 2 dimensional (2D) gel analysis which separates proteins by isoelctric point (isoelctric focusing - IEF) and by monomer mass (SDS-PAGE).The results are shown in Appendix 1B. Although 1D SDS-PAGE appeared encouraging, the 2D gel analysis revealed that the sample was very heterogeneous as multiple spots (each corresponds to a different protein) were observed on the gel. Moreover, most of the protein resolved at ~ 71 kDa which corresponds to the monomer molecular weight of NeTyrA appeared as a smear between pH 4-6. The MS analysis (data not shown) of a tryptic digested sample derived from that region of the gel (Appendix 1B, green box) identified a high content of *E. coli* chaperone proteins (DnaK (pI 5.9) and Hsp90 (pI 5.2)) as well as proteins with affinity to Ni-NTA resin (ArnA (pI 6.9) and GlmS (pI 5.9)). Analysis of the vertical smear observed at pH 8-9 at various protein sizes (Appendix 1B, red box) identified NeTyrA as the primary protein (81% sequence coverage) but

other proteins were detected including ArnA and GlmS. As revealed from the 2D gel analysis, the purified sample contained many *E. coli* proteins and only a small amount of NeTyrA.

2.2.5 Expression of NeTyrA from the synthesized gene and chromatography on Ni-NTA resin

In an effort to improve the expression of NeTyrA in *E. coli*, the *tyrA* gene was optimized for *E. coli* expression and synthesized. Gene synthesis is reported to eliminate factors that can interfere with successful recombinant expression in *E. coli:* A-T richness and AT repeats[(142)], codons that are rarely used in *E. coli* (although we used a Rosetta2 strain to optimize to optimize codon usage), RNA instability motifs[(143)], repeat sequences that yield secondary structures[(144)], and a long hydrophobic N-terminal region within the protein[(145)]. To address the last factor, nucleotide changes were also incorporated to disrupt a hydrophobic stretch of eleven residues at the protein's N-terminus such that amino acids at the second and third position in the primary sequence (valine and methionine) were replaced by four residues (isoleucine, serine, valine and lysine) found in *tyrA* from *A. fulgidus*, the only other trifunctional TyrA characterized to date[(55)].

The synthesized *tyrA* gene was cloned into pET-15b, expressed in *E. coli* and chromatographed on Ni-NTA resin using a step-wise imidazole gradient as described in 2.1.5-2.1.7. A summary of the purification of NeTyrA is presented in Table 2.4 and SDS-PAGE analysis of fractions collected at various stages of the purification are shown in Figure 2.8. SDS-PAGE analysis revealed the presence of a prominent protein band corresponding to ~71 kDa (the predicted mass of monomeric NeTyrA) in the cell lysate, cell-free extract and after Ni-NTA chromatography (Fig 2.7). Additionally, the sample appeared to contain a smaller amount of contaminating proteins compared to NeTyrA derived from the originally cloned gene (Fig 2.5). Similar to our previous observation with the purification of protein from the originally cloned

tyrA gene, protein samples eluting in the flow-through fraction from Ni-NTA chromatography possessed high specific activities suggesting that they contained different forms of catalytically active PD (Table 2.4). The final yield of PD activity units for Ni-NTA purified NeTyrA from a 2 L culture was the same as that obtained previously from a 12 L culture, representing a marked improvement (Table 2.4). Additionally, the protein showed a fivefold increase in specific activity compared to our previous preparations.

Step	Volume (mL)	Total Protein (mg)	Total PD Activity (U)	PD S. A. (U/mg)	Yield PD (%)	Purification fold
Cell free extract	55	330	361	1.1	100	1.0
Flow through	55	225	103	0.5	29	0.4
Pooled & concentrated	65	6.5	42	6.5	12	5.9

Table 2.4: Purification table of NeTyrA. Protein expressed from a 2 L cell culture derived from the synthesized gene. Protein expression was carried out overnight at 18°C after induction with 0.4 mM IPTG. Prephenate dehydrogenase activity determined at 80°C: 0.5 mM prephenate, 2 mM NAD^+, 50 mM Hepes, 200 mM NaCl, pH 8. Non enzymatic rate subtracted. CM and PDT activities were also determined with 0.5 mM substrate, yielding 2.7 and 6.8 U/mg, respectively.

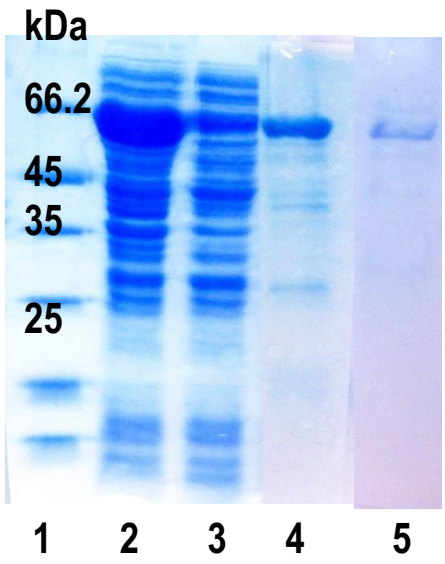


Figure 2.7: SDS-PAGE analysis of NeTyrA purification by Ni-NTA chromatography. Proteins were expressed from the synthesized gene. Lane 1: Molecular Weight Marker, Lane 2: Cell-lysate, Lane 3: Cell-free extract, Lane 4: Pooled protein from Ni-NTA chromatography, Lane 5: Thrombin-treated and reapplied on Ni-NTA resin (untagged).

2.2.6 Mass spectrometry analysis of NeTyrA and IhTyrA

2.2.6.1 Intact protein mass determination using ESI-MS

The presence of full-length archaeal TyrA proteins in preparations following Ni-NTA chromatography were confirmed using mass spectrometry. We tried several methods of sample preparation: a method developed by Kaback *et al*[146] for the analysis of membrane proteins, which had been successfully used in our lab for intact mass analysis of *E. coli* and *A. aeolicus* TyrAs[24]; the zip-tip purification method that was successfully employed for MS analysis of *H. influenzae* TyrA[95]; and by HPLC-MS outlined in section 2.1.15. Regrettably, we were not able to detect the intact mass of NeTyrA that was expressed from the cloned genomic *tyrA* by any of the methods outlined above. However, we succeeded in obtaining the intact mass of NeTyrA that was expressed from the synthesized gene as well as that of IhTyrA using the HPLC-MS method.

ESI-MS analysis (Fig 2.8) revealed that both His-tagged NeTyrA and IhTyrA proteins were present as the major species but at a mass smaller than predicted. The spectrum of NeTyrA (Fig 2.8 A) showed one major peak at $[M+H^+]$ of 71 353 Da which is 128 Da lower than the expected mass of 71 481 Da. Similarly, the spectrum of His-tagged IhTyrA (Figure 2.8, B) showed one major peak at 42 330 Da which is 134 Da lower than the expected mass of 42 364 Da. We attributed the mass difference to a post-translational removal of first methionine at the tag's N-terminus. Additional peaks of 35 675 Da and 23 785 Da in the mass spectrum of NeTyrA correspond to doubly and triply charged species, respectively of the same protein. ESI-MS also detected a lower abundance peak in the NeTyrA sample with $[M+H^+]$ of 71 532 corresponding to a mass addition of 179 Da relative to the major species (Figure 2.8, A), and lower abundance peaks in the IhTyrA sample with $[M+H^+]$ of 42 409 and 42 487, which represent mass additions

of 179 Da and 257 Da, respectively (Figure 2.8, B). We attribute these species to post-translational α-N-gluconoylation α-N-6 phosphogluconoylation of the hexa-His tag as explained by Geoghegan et al [(147)]. These adducts on IhTyrA were removed by reapplication of the thrombin-treated sample to Ni-NTA resin and generated untagged protein of 40 612 Da (40 613 Da expected) (Fig 2.8 C). A similar strategy to remove the affinity tag from NeTyrA, however, resulted in a very low yield of the trifunctional enzyme with reduced activity, which prevented determination of its intact mass. ESI-MS analysis (Fig 2.8 B and C) also detected a lower abundance, smaller molecular weight species (~ 30 kDa) in His-tagged and untagged samples of IhTyrA, which was also observed by SDS-PAGE analysis (Fig 2.6, lane 7 and 8). Identification of this degradation product is further discussed in 2.2.6.2 and in Chapters 3 and 4.

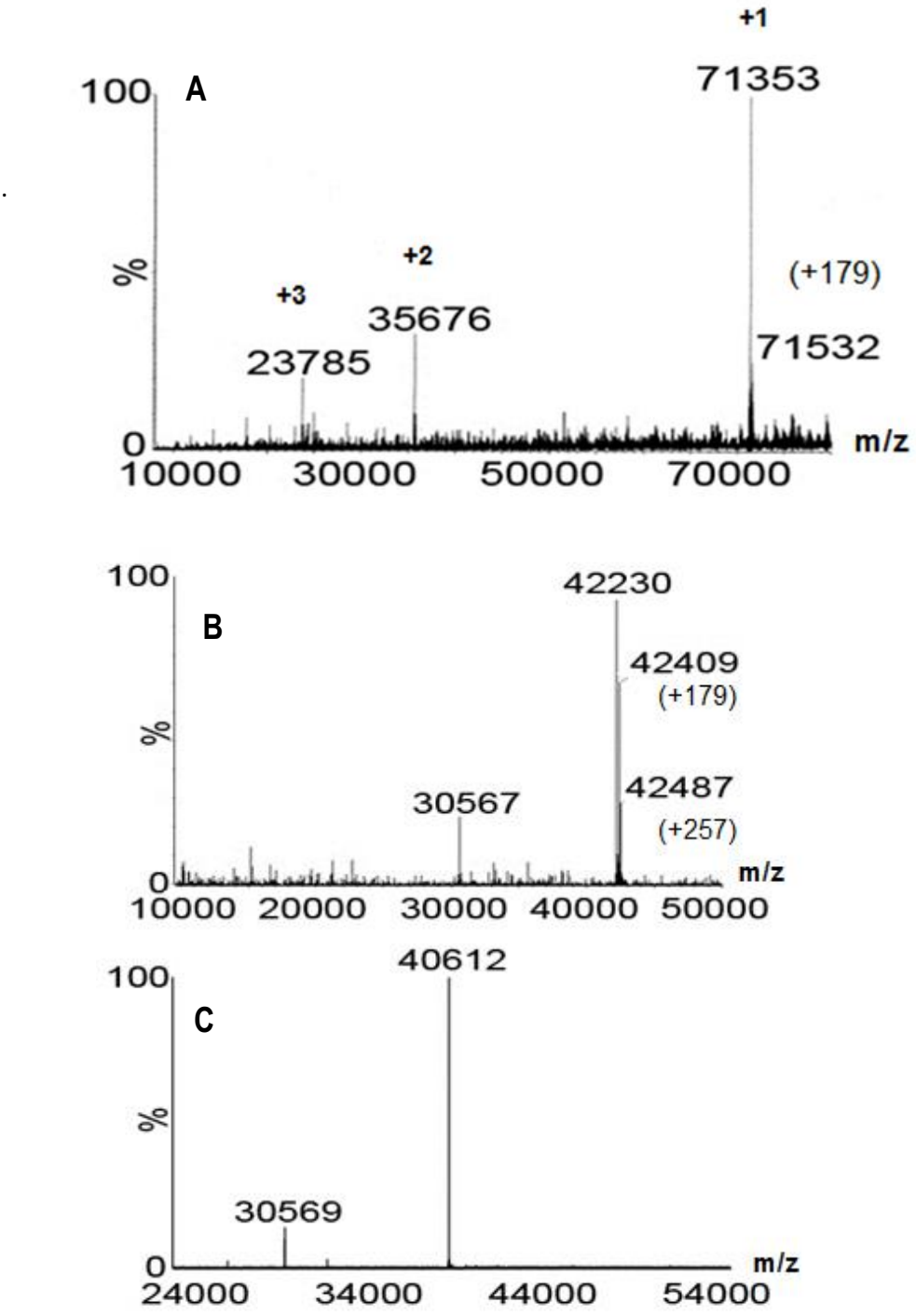


Figure 2.8: **ESI-MS analysis of NeTyrA and IhTyrA.** (A) His-tagged NeTyrA, (B) His-tagged IhTyrA and (C) IhTyrA, tag removed. Samples were analyzed by LC-MS as described in Section 2.1.15. Mass additions correlated with α-N-gluconoylation and α-N-6-phosphogluconoylation of the His tag are highlighted in parentheses. In (A) singly, doubly and triply charged species and depicted as +1, +2 and +3, respectively. Peaks assigned in (B) and (C) represent singly charged species. The Y-axis represent % of abundance relative to the major species.

2.2.6.2 Tandem mass spectrometry analysis

In order to further evaluate the homogeneity of the NeTyrA protein sample, we performed in-gel tryptic digestion of the ~ 71 kDa band from the denaturing gel of purified NeTyrA followed by tandem MS analysis. The analysis detected peptides from all three domains (PD, CM and PDT), which yielded a total protein sequence coverage of 76%. The results are summarized in Table 2.5 and in Appendix 2A (Appendix 2A tabulates the calculated masses of the peptides from *in-silico* tryptic digestion of NeTyrA along with the peptides in Table 2.5 observed by tandem MS analysis). The Mascot score (a "confidence score") is used in MS analysis to rank proteins according to the total amount of evidence supporting the identification of each protein. NeTyrA was ranked with the highest Mascot score, along with a variety of lower ranking *E. coli* proteins (Table 2.5). Among the proteins which co-purified with NeTyrA are chaperones DnaK and Hsp90 which assist in proper enzyme folding, as well as ArnA and GlmS proteins which are reported to possess affinity to nickel (Table 2.5) [(148)]. Given the sensitivity of mass spectrometry it is surprising that ESI-MS did not detect any of these *E. coli* contaminants in the NeTyrA sample. These *E. coli* proteins were detected by tandem MS analysis in all preparations of NeTyrA expressed from the cloned genomic DNA (see representative analysis in Appendix 1A), although in much higher abundance relative to TyrA.

gene	Protein name (~ 71 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
NEQ192* of *N. equitans	**PD-CM-PDT**	**Q74NC4**	**1456**	**55**	**76**	**71.35**
glmS	**L-gln-D-fru-6-p aminotransferase**	**P17169**	**568**	**29**	**64**	**67**
dnaK	**Chaperone DnaK**	**P0A6Y8**	**301**	**24**	**48**	**69**
ppiD	**Peptidyl-prolyl cis-trans isomerase D**	**P0ADY1**	**88**	**13**	**28**	**68**
htpG	**Chaperone Hsp90**	**P0A6Z3**	**46**	**7**	**16**	**71.42**
arnA	**Polymyxin resistance protein**	**P77398**	**42**	**7**	**24**	**75**
rpsA	**30S ribosomal protein**	**P0AG67**	**38**	**6**	**16**	**61**
sdhA	**Succinate dehydrogenase**	**P0AC41**	**29**	**5**	**10**	**64**

Table 2.5: Orbitrap Velos MS analysis of a tryptic digestion of the 71 kDa band sample from Ni-NTA purified NeTyrA. Listed are parameters for NeTyrA and *E. coli* proteins whose peptides were detected by MS.

2D gel analysis was performed on the NeTyrA sample that was analyzed by SDS-PAGE (Fig 2.7) and by tandem MS (Table 2.5). The results shown in Fig 2.9 revealed a number of protein bands highlighting the heterogeneity of the sample. Tandem MS analysis of tryptic peptides derived from the most prominent bands on the gel (green boxes) identified chaperone proteins DnaK, Hsp90 and GroEL (boxes 3 and 4), SlyD (box 5), and proteins with affinity to Ni-NTA, ArnA, GlmS (boxes 3 and 4) and the lower molecular weight (~ 16 kDa) Nur- ferric uptake regulator (box 6). Tandem MS analysis of the bands in the pH range of 8-9 (red boxes 1, 2) (predicted pI 8.8 of NeTyrA), successfully identified the trifunctional enzyme with the highest Mascot score although *E. coli* proteins such as chaperones PpiD and GroEL were also identified. Additionally, peptides from NeTyrA were detected in the band with molecular weight of ~ 18 kDa (box 7) suggesting the presence of either a degradation product or a truncated version of TyrA. Overall, the 2D gel profiles of the NeTyrA sample derived from the synthesized gene and

from the cloned genomic DNA (Appendix 2A) are comparable in their interpretations. In both samples the majority of the proteins were of *E. coli* origin (chaperones and proteins with affinity to nickel) with only a modest amount was NeTyrA. Moreover, both samples contained smaller sized fragments of NeTyrA suggesting possible degradation products due to the instability of the enzyme when separated from chaperones. We attempted to further purify NeTyrA by cleaving the His tag with thrombin and reapplying the sample onto Ni-NTA resin (see SDS-PAGE analysis in Fig 2.7, Lane 5 (untagged) and Lane 4 (His tagged protein prior to thrombin treatment)). As mentioned previously, this strategy was not particularly successful since it resulted in a very low yield of the trifunctional enzyme. Moreover, as revealed by tandem MS analysis, the untagged NeTyrA sample was purified only from contaminating *E. coli* proteins with affinity to Ni-NTA affinity resin but retained the chaperone proteins (data not shown).

Although not homogeneous, the His-tagged NeTyrA protein sample exhibited significant enzyme activity. Accordingly, we performed the kinetic analysis and some biophysical characterization with the His-tagged NeTyrA sample.

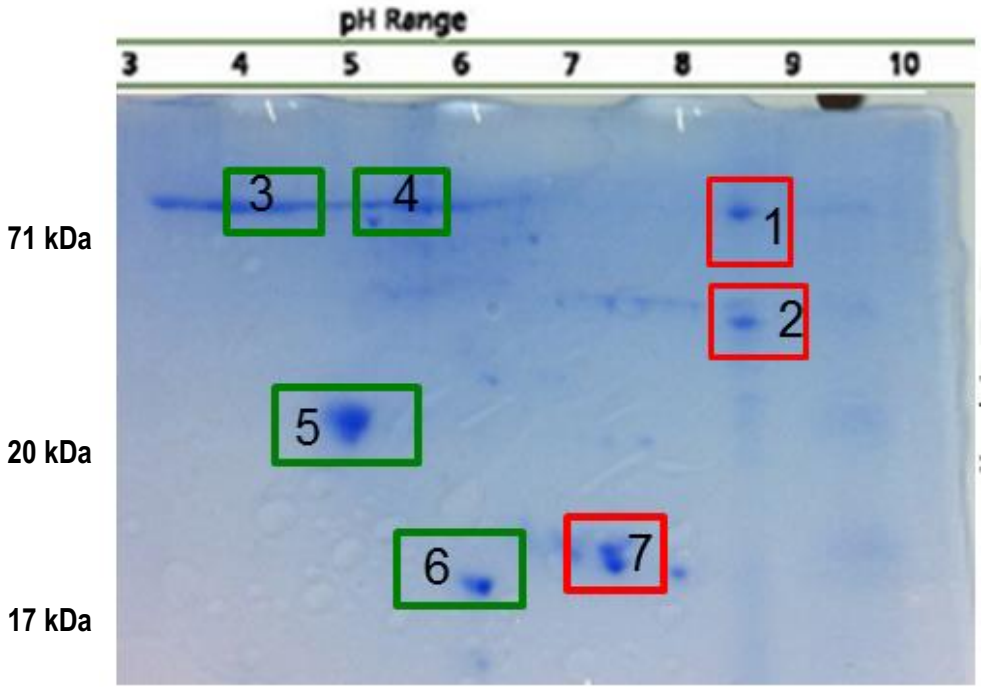
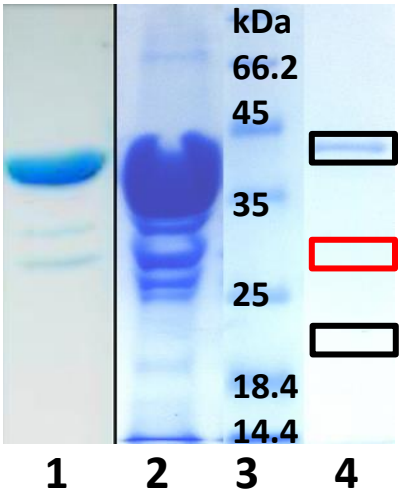


Figure 2.9: 2D gel analysis of NeTyrA expressed from the synthesized gene and purified by Ni-NTA chromatography. Protein sample (170 µg) was analyzed. Green and red boxes highlight samples which were subjected to in-gel tryptic digestion followed by tandem MS analysis. Red boxes: analysis ranked NeTyrA with the highest Mascot score. Green boxes: analysis ranked *E. coli* proteins with the highest Mascot score.

Tandem mass spectrometry was applied to further characterize IhTyrA, in particular to identify the ~ 30 kDa species observed by SDS-PAGE (Fig 2.6) and by ESI-MS analysis of His-tagged and untagged TyrA samples (Fig 2.8, B and C). We performed in-gel tryptic digestion of the ~ 30 kDa fragment and the ~ 41 kDa band (corresponding to a full-length, untagged TyrA protein) from the denaturing gel of IhTyrA sample followed by tandem MS analysis. Appendix B2 summarizes the calculated masses of the peptides from *in-silico* digestion of *I. hospitalis* TyrA along with the peptides observed by Orbitrap Velos MS analysis from tryptic digestions of both protein bands. While the tandem MS analysis of the ~ 41 kDa band detected peptides from both

CM and PD domains (sequence coverage 61%) , only peptides from the C-terminal portion of IhTyrA were detected in the ~ 30 kDa fragment (See Fig 2.10 for sequence coverage of the ~ 30 kDa fragment). Peptides from the N-terminal portion of the protein that comprised the first 80 amino acids were not observed by tandem MS analysis of the ~ 30 kDa fragment. Thus, we attribute this species to a degradation product comprising the C-terminal portion of the protein.



GSHMSENPLESLKVKRRELDKLDKEILRLLKRRFEIVKEITDTKKNLGLPVYDRDREEEVMVTRTVWGLELGIPQEF
TKEMFKMILEESKKIQLYTPEKVYVGIYGYGGMGEQLVKVFSRAGHRVVVTGRNLEKAEGLAKRFKVEWGEPKEVAK
EVEWLILAVPPKAVPGLVKELAPLMRSGALLSDISSVKKTLVEEVLKVLPEYIEYISLHPLFGPEVEPLGETVVVVP
VKSYDYWVRLVQNIFVSMGFEVITSTPEEHDRAMAVTQVLHHFALVSLDEAAKKLSKEYGVDYMRYATRSFKKTLET
IQRLKELSEVIDEIQEMNEYAAHAREEFLKVASQMDKRWRKG

Figure 2.10: SDS-PAGE analysis and sequence coverage by tandem MS analysis of IhTyrA 30 kDa fragment. Lane 1: IhTyrA purified by Ni-NTA chromatography, tag removed (7 µg), Lane 2: Concentrated IhTyrA (64 µg), Lane 3: Molecular Weight Marker, Lane 4: 80-fold dilution of sample in Lane 2, bands from this sample (highlighted in boxes were subjected to in-gel tryptic digestion and analyzed by tandem MS. Red box highlights band corresponding to ~ 30 kDa fragment while black boxes highlight the band corresponding to full-length IhTyrA and the region of the gel that used for the blank sample. IhTyrA sequence highlighted in gray represents the region from which the peptides were identified by MS in sample that contained tryptic digestions of IhTyrA ~ 30 kDa band.

2.2.7 Native molecular weight determination of NeTyrA and IhTyrA

Analysis of IhTyrA by size exclusion FPLC showed that the majority of the protein eluted at a molecular weight of 69 kDa with a lower abundance species at 151 kDa (Fig 2.11). This is consistent with the resolution of dimeric and tetrameric forms of the enzyme. The dimeric fold is in agreement with the quaternary structures of other known TyrA proteins [40, 51, 52, 91, 92, 149].

Interestingly, both of these fractions contained protein that was catalytically active with PD activity. A small amount of protein eluted at ~7 KDa but this did not possess PD activity. IhTyrA was analyzed further by analytical ultracentrifugation sedimentation velocity experiments at high and low protein concentrations. The results were in agreement with those obtained from SEC-FPLC showing mainly a dimeric species (85 kDa) in the presence of a lower MW abundance tetramer species (170 kDa) (Fig 2.12). A value of 85 kDa obtained from these sedimentation velocity experiments more closely agreed with the calculated dimer mass of 84.5 kDa. Analysis also revealed the presence of a protein in low abundance with a sedimentation coefficient >11 consistent with a large aggregate. Additionally, one or more species with a sedimentation coefficient <3 corresponding to molecular weight of ~ 20 kDa or less were also seen. We attempted to determine if the dimeric and tetrameric species were in dynamic equilibrium at ambient temperature. Dimeric *E. coli* CM-PD was previously reported to convert partially to a tetramer upon binding NAD^+ or more fully with NAD^+ and L-Tyr as a form of end product regulation [(98)]. However, size exclusion chromatography performed in the presence of 2 mM NAD^+ (the enzyme was not inhibited by L-Tyr as discussed in Chapter 3) did not appear to shift the relative abundance of these two forms of IhTyrA from the dimer towards the tetramer (data not shown). Additionally, the molecular weight of the major species was not altered with a tenfold increase in protein concentration as illustrated by sedimentation velocity experiments (Fig 2.12) although the proportion of the tetramer appeared to increase at the expense of the species exhibiting the lowest sedimentation coefficient.

Size exclusion chromatography of a sample of Ni-NTA-purified NeTyrA yielded variable results depending on the preparation, most certainly due to the heterogeneity of the sample. The sample was therefore heat-treated (1 h, 95°C) after chromatography on the Ni-NTA resin. Heat treatment removed many of the *E. coli* proteins although the fraction did retain PD activity.

As expected, the low yield of protein produced a weak and rather complex absorbance spectrum at 280 nm. Nevertheless, size exclusion chromatography of this sample showed that PD activity could be detected only in a fraction with an elution volume of 12 mL (Fig 2.11). This volume corresponded to a native molecular weight of ~ 533 000 Da. Given a monomer molecular weigth of 71 kDa, this large size could be attributed to either an octamer or a heteromeric complex with *E. coli* chaperone proteins. The chromatography was repeated in the presence of 1 mM of L-Phe since this ligand is reported to induce the formation of higher order species in PDT enzymes[(102)]. However, no significant shift in molecular weight was observed.

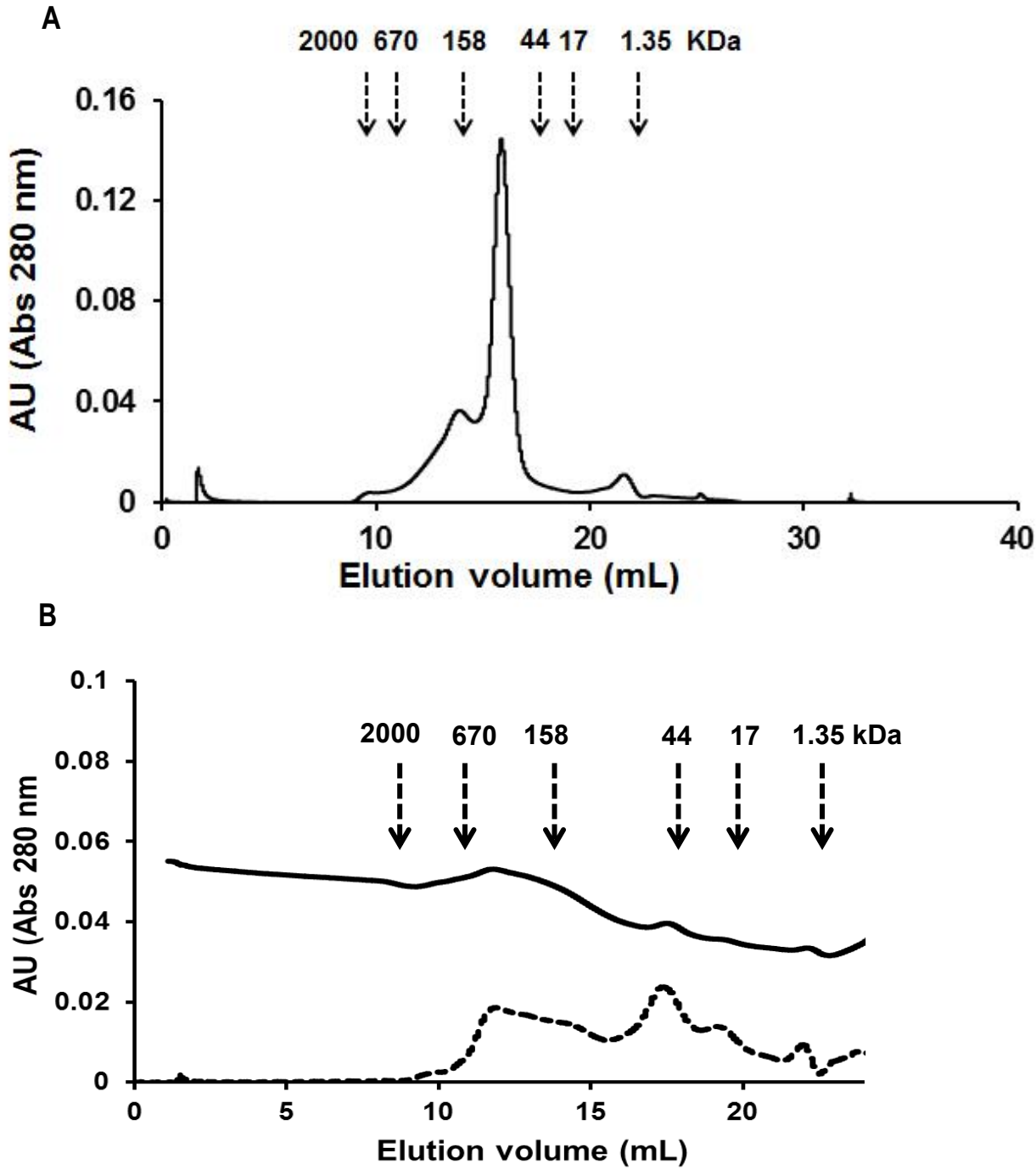


Figure 2.11: Size exclusion-FPLC analysis of purified TyrA proteins. (A) IhTyrA and (B) NeTyrA. Arrows show the elution pattern of each of the indicated molecular weight standards (BioRad). (A) IhTyrA showing PD activity eluted in fractions between 13-17.5 mL. The major peak is at 15.9 mL with a calculated molecular weight of 69 kDa, while a second active peak at 13.9 mL corresponds to 161 kDa.(B) NeTyrA was heat-treated (1 h, 95°C) prior to analysis by SEC-FPLC. The active TyrA protein eluted in a fraction at 12 mL, and corresponded to a calculated molecular weight of 533 kDa. Two other peaks associated with protein (but not exhibiting PD activity) gave calculated molecular weights of 58 kDa and 7 kDa.

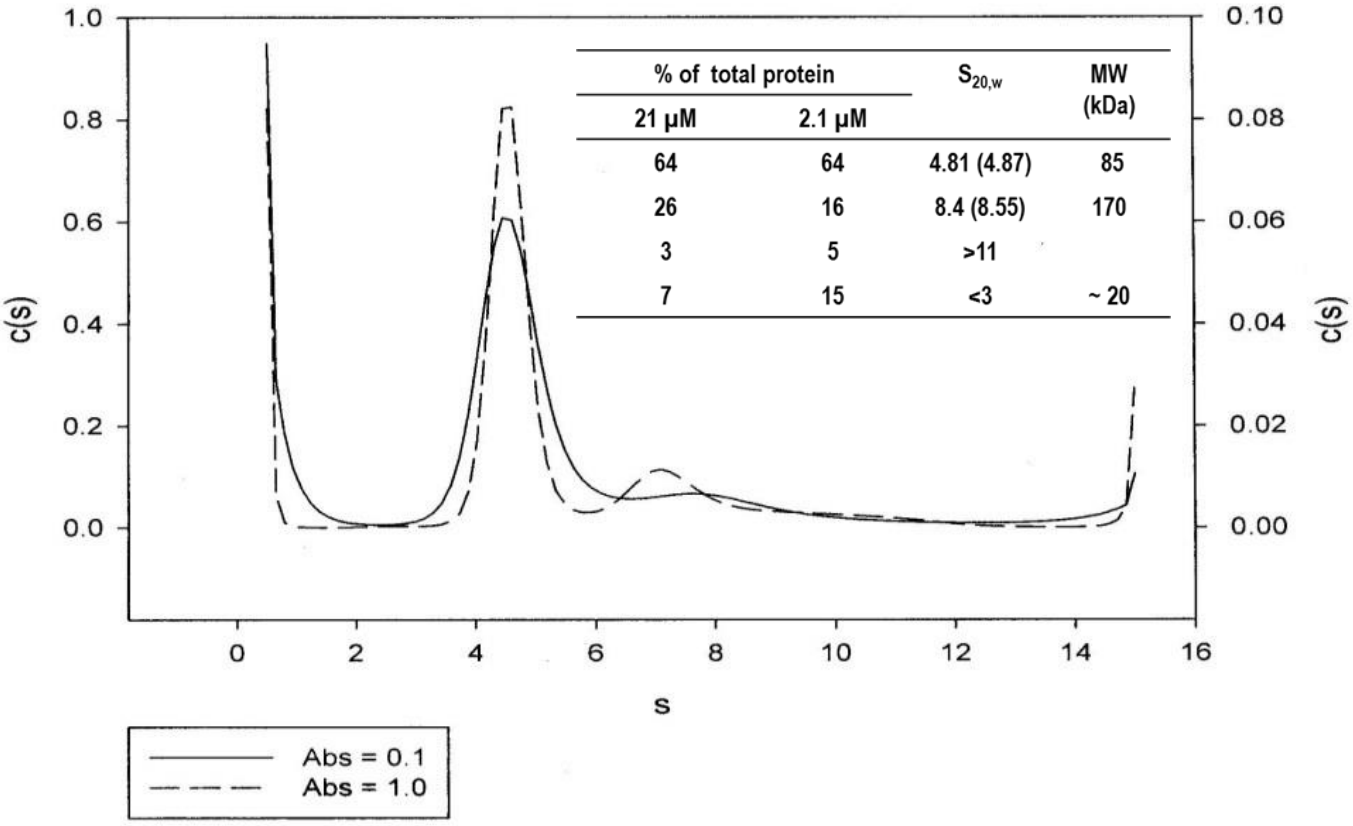


Figure 2.13: Analytical ultracentrifugation analysis of IhTyrA. Protein samples were at monomer concentrations of 21 µM (solid line) and 2.1 µM (dotted line), corresponding to absorbance 280 nm reading of 1 and 0.1, respectively. Values for $S_{20,W}$ and molecular weight were determined as described in materials and methods. $S_{20,W}$ values obtained for 2.1 µM *I. hospitalis* CM-PD are indicated in parentheses. As a control, *E. coli* CM-PD (0.15 mg/mL) was analyzed by a sedimentation velocity experiment and the observed peak corresponded to a molecular weight of 72 kDa (data not shown).

2.3 Discussion

This chapter presents the first report of the expression, partial purification and functional analysis of recombinantly expressed TyrA proteins from the symbiotic archaea *N. equitans and I. hospitalis.* We also present our efforts to characterize the protein preparations by mass spectrometry analysis and to determine the protein's native molecular

weight. Our findings on the characterization of the archaeal TyrAs, which are presented in this Chapter and in Chapter 3, are now published in *Extremophiles*[150].

The *tyrA* genes from the only known archaeal parasite *N. equitans* and that of its host *I. hospitalis* are of particular interest; the former is predicted to encode an unusual protein encompassing three activities PD-CM-PDT, while the latter, a bifunctional CM-PD. In most microorganisms, TyrA enzymes exist as monofunctional or bifunctional fusions of two activities, CM-PD. To date, the three domain assembly is evident in TyrA proteins from only two other organisms: PD-CM-PDT of the archaeon *A. fulgidus* and the putative CM-PDT-PD of the algae *E. siliculosus*[54, 55].

Heterologous expression and purification by Ni-NTA chromatography yields active but not homogeneous TyrA enzymes

The *tyrA* genes from *N. equitans* and *I. hospitalis* were cloned into a pET-15b vector, the proteins were heterologously expressed in *E. coli* with a 20-residue N-terminal extension that included a hexa-His tag to facilitate purification by Ni-NTA affinity chromatography and a thrombin recognition site to allow removal of the tag. Affinity purification of IhTyrA was reasonably effective as judged by SDS-PAGE analysis (Fig 2.6, Lane 7). The final yield of PD activity was approximately 26% for IhTyrA, which is comparable to that of the hyperthermophilic monofunctional PD from *A. aeolicus*[52]. Western blot and SDS-PAGE analyses that monitored the NeTyrA protein during its expression and purification, however, revealed only a modest yield of soluble protein (Fig 2.3-2.5). Moreover tandem MS and 2D gel analysis revealed that the purified NeTyrA sample was contaminated with multiple *E. coli* proteins, such as chaperones to assist in the folding of newly synthesized proteins, and proteins with affinity to nickel (Fig 2.9 and Table 2.5). Nevertheless, we performed kinetic studies on the archaeal TyrA enzymes obtained after

Ni-NTA affinity chromatography and confirmed the presence of all predicted enzymatic activities CM, PD and PDT for the trifunctional enzyme and CM and PD for the bifunctional TyrA and at assay temperature (80°C) close to the temperature optimum for the growth of *I. hospitalis* and *N. equitans*.

We attempted to improve the expression and purification of NeTyrA in order to achieve higher yields of soluble, homogeneous protein. This included cloning the gene into other expression vectors, altering expression conditions (reducing the induction temperature and concentration of IPTG) and adding extra chromatography steps (based on ion exchange, hydrophobic interactions, size exclusion or $NADP^+$ - affinity). However, none of these modifications were particularly effective in increasing the yield of soluble enzyme with significantly higher specific activity. Synthesis of the *tyrA* gene, however, yielded more encouraging results.

While there are many factors that can interfere with heterologous protein expression in *E. coli* and are routinely eliminated by gene synthesis we thought to reduce the A-T content of *tyrA* gene (from 68% to 55%) and to reduce the hydrophobicity at the NeTyrA's immediate N-terminus by replacing it with a less hydrophobic sequence found in *tyrA* from the archaeon *A. fulgidus* which encodes a PD-CM-PDT[(55)]. The final yield of NeTyrA PD activity was approximately 12%, which is similar to that obtained for the trifunctional TyrA enzyme from *A. fulgidus*[(55)] and represents a sixfold increase in the active enzyme per L of culture over that NeTyrA previously obtained from the cloned NeTyrA gene.

Mass spectrometry confirms the presence of the archaeal TyrA proteins

Purification of NeTyrA expressed from the synthesized gene (but not from the cloned gene) allowed intact ESI-MS analysis of the protein. ESI-MS analysis (Fig 2.8) confirmed the

presence of NeTyrA and IhTyrA although both His-tagged proteins lacked the first methionine at the tag's N-terminus. Observed masses of 71 353 Da (NeTyrA) and 42 230 Da (IhTyrA) are in good agreement with the expected values of 71 350 and 42 233 Da. Post-translational modification where the N-terminal Met is hydrolyzed, is frequently reported in many proteins including recombinant *E. coli* CM-PD[(94)]. ESI-MS also detected lower abundance species with mass additions of 178 Da and 258 Da, which are likely due to post translational α-N-gluconoylation and α-N-6 phosphogluconoylation of the hexa-His tag. α-N-gluconolacton and α-N-6 phosphogluconolacton are intermediates in the pentose phosphate pathway. High level expression of recombinant proteins in *E. coli* BL21(DE3) can interfere with this pathway resulting in accumulation of these gluconolactones which are potent agents for covalent modification of proteins, notably those that are His-tagged[(147)]. Covalent glycation of His tagged proteins expressed in BL21(DE3) is reported to reduce activity or to interfere with crystallization. These gluconolacton adducts were readily removed by reapplication of the thrombin-treated sample to Ni-NTA resin and generated an untagged IhTyrA (Fig 2.8, C). A similar strategy to remove the affinity tag from NeTyrA, however, resulted in a very low yield of the trifunctional enzyme with reduced activity, which prevented determination of its intact mass.

ESI-MS and SDS-PAGE analysis (Fig 2.6 and Fig 2.8, B and C) also detected a lower abundance, smaller molecular weight species (~ 30 kDa) in both His-tagged and thrombin-treated samples of IhTyrA. As one of our longer term goals is to crystallize this bifunctional TyrA, we sought to understand the origin of this fragment. Interestingly, Western Blotting with Anti-His antibodies of the cell-free extract (Fig 2.4) and the retention of the fragment on Ni-NTA affinity resin suggested that this fragment belonged to the N-terminal portion of the protein. However, tandem MS analysis of tryptic digestions of protein originated from the ~ 30 kDa band showed unequivocally that the IhTyrA fragment was made of residues within the C-terminal portion of the

protein, which constitutes the PD domain. The 80 amino acids from the N-terminal portion of the protein comprising the mutase domain were absent in this fragment. It is likely that the 30 kDa fragment is due to cleavage between Lys80 and Met81 at the C-terminal end of the mutase domain catalyzed by non-specific serine proteases in the *E. coli* cell[(151)]. N-terminal degradation of TyrA was previously reported during the heterologous expression and purification of a monofunctional PD from the hyperthermophilic bacterium *A. aeolicus*[(89)]. In the case of *A. aeolicus* PD, proteins could be obtained that were missing 19 and up to 23 residues from PD's N-terminus. Moreover, the presence of the two species in the protein sample interfered with its crystallization and only the independently expressed Δ19PD TyrA yielded diffraction quality crystals[(89)].

Although SDS-PAGE and ESI-MS analysis of trifunctional NeTyrA revealed a relatively homogeneous preparation (Fig 2.7 and 2.8, A), tandem mass spectrometry of purified NeTyrA identified a number of co-eluting *E. coli* proteins, some with a similar monomer size to the target enzyme (~ 71 kDa). Among the co-purified proteins were chaperones DnaK and Hsp90 which assist in proper enzyme folding, as well as ArnA and GlmS proteins with affinity to nickel (Table 2.5)[(148)]. Moreover, the 2D gel analysis (Fig 2.9) revealed that NeTyrA comprised only a very modest fraction of the total protein purified by Ni-NTA chromatography. Heat treatment at 95°C of the Ni-NTA purified sample was helpful in eliminating many *E. coli* proteins and hence would be a useful step in our purification strategy. We show in Chapter 3, however, that this heat treatment was not effective in the removal of chaperone proteins.

According to Mehlin *et al*[(142)] a variety of factors that could make the heterologous expression of soluble protein in *E. coli* very challenging may apply to NeTyrA. Some of these factors cannot be overcome by *tyrA* gene synthesis. Among these are a large monomer molecular weight (>65 kDa), basic pI, greater protein disorder and a low hypothetical annotation

to *E. coli* proteins. Indeed, NeTyrA meets some of these criteria with its large monomer molecular weight (~ 71 kDa), pI of 8.8 and low sequence similarity to *E. coli* CM-PD or CM-PDT. Moreover, bioinformatics programs ProteinPredict and PSIPRED reveal that NeTyrA is predicted to retain a stretch of 17 disordered residues between positions 240-257 (at the beginning of the CM domain of the protein). Thus, all of these factors likely contributed to the poor expression of soluble NeTyrA.

Native molecular weight determination establishes higher order structure of archaeal TyrA proteins

Both size exclusion FPLC and analytical ultracentrifugation sedimentation velocity revealed a native molecular weight of IhTyrA (~ 70-85 kDa) suggesting a dimeric structure (Figs 2.11 and 2.12). The dimeric fold is in agreement with the quaternary structures established for most purified TyrA proteins characterized to date including bifunctional CM-PD from *E. coli* and *H. influenzae,* monofunctional PD from *A. aeolicus and* monofunctional cyclohexadienyl dehydrogenase from *Z. mobilis*[40, 51, 52, 91, 92, 149]. Moreover, despite the poor primary sequence identity, the available crystal structures of PD from *A. aeolicus, S. mutans and H. influenzae* as well as AD from *Synechocystis* sp. PCC 6803 revealed a common architecture: an N-terminal nucleotide binding domain, C-terminal dimerization domain and an active site, located at the interface between the two domains comprised of residues that are shared between adjacent monomers (Chapter 1, Fig 1.9 and 1.10) [41, 90-92]. It seems that the global dimeric fold of *I. hospitalis* PD is similar to that established by the available crystal structures. A model of *I. hospitalis* PD based on crystal structures of *A. aeolicus* and *H. influenzae* PDs is shown in Appendix 7B, which supports dimerization of the monomers at the C-terminal domain.

Size exclusion FPLC and sedimentation velocity analysis of IhTyrA revealed an additional lower abundance, catalytically active species corresponding to molecular weights (160 -170 kDa) (Fig 2.11 and 2.12), which we attributed to a tetrameric form of the enzyme. However, these forms do not appear interconvertible at ambient temperatures. It was previously reported by Hudson *et al*[98] and later by Bonvin *et al*[52] that dimeric *E. coli* CM-PD converts partially to a tetramer upon binding of NAD^+ or more fully with NAD^+ and L-Tyr as a form of end product regulation. However, dimeric IhTyrA did not appear to undergo this interconversion when size exclusion chromatography was performed in the presence of 2 mM NAD^+. It has been reported that proteins *in vitro* may undergo concentration-dependent oligomerization. For example, dimeric enzyme EntA, from the enterobactin biosynthetic pathway of *E. coli*, is reported to form a tetramer at high protein concentrations [152]. However sedimentation velocity analysis of IhTyrA revealed that the molecular weight of the major species (which is correlated to a dimer) was not altered with a 10-fold increase in protein concentration (Fig 2.12). Additionally, it may be that at ambient temperatures, well below the optimum of the enzyme's function, TyrA proteins from hyperthermophilic organisms do not possess the correct conformation to undergo a quaternary structural change.

In contrast with bifunctional IhTyrA that appeared to adopt a dimeric native fold, size exclusion chromatography of a heat-treated NeTyrA sample revealed a catalytically active oligomer with a relatively high apparent molecular weight of $\sim$ 533 kDa. We attribute this unusually large size to either an octamer or a heteromeric complex with *E. coli* chaperone proteins. Interestingly, the only other known archaeal trifunctional PD, from *A. fulgidus*, is also reported to possess an atypically large native assembly by FPLC analysis, in that case a hexamer [55]. Higher oligomerization states of cytoplasmic enzymes observed in hyperthermophilic prokaryotes are thought to contribute to their heat stability [153]. For

example, the octameric acetyl-CoA synthetases from the hyperthermophilic *I. hospitalis* and *Pyrobaculum aerophilum* are unique in their stoichiometry and significantly distinct in their oligomeric state from those of mesophilic organisms that are usually monomeric or dimeric[(154)]. Thus, it is likely that both trifunctional TyrA proteins from the hyperthermophilic *N. equitans* and *A. fulgidus* may have acquired the high oligomeric fold through the process of evolutionary adaptation to the extremely hot temperature of their exogeneous environment.

As a form of allosteric control, PDTs are reported to undergo an L-Phe-induced formation of higher order species [(102)]. For example monofunctional PDT from *B. subtilis*[(155)] and the bifunctional CM-PDT from *Salmonella typhymurium*[(156)] are reported to dimerize upon the binding of L-Phe, while dimeric *E. coli* CM-PDT is reported to form a tetramer in the presence of its feedback inhibitor[(102)]. However, no significant shift in molecular weight of NeTyrA (the trifunctional PDT) was observed in the presence of 1 mM L-Phe (inhibition studies by L-Phe are discussed in Chapter 3). It has been proposed by Zhang and colleagues [(102)] that the N-terminal CM domain induces the oligomerization of *E. coli* CM-PDT since the individually expressed PD domain did not yield a tetramer in the presence of L-Phe. It is possible that L-Phe-induced change in NeTyrA's molecular weight was not observed due to the heterogeneity of the sample, that is, the NeTyrA preparation contained chaperones that might interact with the CM domain and prevent any ligand-induced changes in shape. Alternatively, it may be that NeTyrA is similar to the PDT from *Methanocaldococcus jannaschi*, whose L-Phe induced conformational change is not accompanied with protein oligomerization[(107)].

2.4 Summary

TyrA proteins from hyperthermophilic archaeal symbionts *N. equitans and I. hospitalis* were expressed as His-tagged proteins in *E. coli* and were chromatographed on Ni-NTA resin.

This approach yielded substantial levels of active IhTyrA suitable for all kinetic and biophysical studies. The expression of soluble *N. equitans* TyrA was relatively poor, however, but this could be improved by producing the protein from a synthesized gene. We demonstrated for the first time activity of the archaeal TyrA proteins which correlates with the function of the predicted domains. ESI-MS analysis confirmed the masses of the proteins and identified a small abundance species of ~ 30 kDa in *I. hospitalis* TyrA sample which we attribute to a degradation product from the C-terminal portion of the protein. Tandem mass spectrometry analysis confirmed the presence of all enzymatic domains of the purified TyrA proteins and identified *E. coli* proteins in preparations of NeTyrA sample including chaperones, which co-express with the archaeal protein to assist in its proper folding. We also evaluated the native molecular weight of the archaeal TyrA proteins using analytical size-exclusion chromatography and sedimentation velocity analyses. The native MW of IhTyrA appeared to be dimeric, similarly to other characterized TyrA proteins. Interestingly, the trifunctional PD-CM-PDT from *N. equitans* presented an unusally large assembly suggesting an octamer. The oligomeric assembly of IhTyrA and NeTyrA was not changed in presence of NAD^+ and L-Phe, respectively, although, these ligands were reported to induce a shift in the oligomeric states of PD and PDT, respectively, from other prokaryotes. We are the first to characterize these archaeal TyrA proteins that likely represent uncommon members of the TyrA protein family.

In the next chapter, the effects of temperature, salt and pH on enzyme function are explored as well as the effect of temperature on structure. We also determine the kinetic parameters for all enzymatic reactions catalyzed by the TyrA proteins and determine the effects of the end products L-Tyr and L-Phe on enzyme activity. The stability of IhTyrA towards proteolytic cleavage is also evaluated and we show how this leads to the identification of a stable dehydrogenase fragment.

Chapter 3

Protein Stability and Kinetic Properties of TyrA Proteins from *N. equitans* and *I. hospitalis*

3.0 Introduction

TyrA family dehydrogenases exhibit varied specificity to their substrates prephenate and L-arogenate, and to the pyridine nucleotide cofactors NAD^+ and $NADP^+$ [29]. It is commonly observed, that prephenate dehydrogenases are NAD^+ -dependent, while arogenate dehydrogenases generally use $NADP^+$. TyrA enzymes studied by Turnbull and colleagues have centered mainly on NAD^+-dependent prephenate dehydrogenases such as from the mesophilic organisms *E. coli* (CM-PD) and *H. influenzae* (CM-PD) and from the thermophile *A. aeolicus* (PD)[24, 57, 95]. Interestingly, we have discovered through amino acid sequence alignment of a number of TyrA proteins with PD domains of *I. hospitalis and N. equitans* that both symbiotic archaea possess *tyrA* genes that may give rise to $NADP^+$-specific PD activity (Fig 1.8). Moreover, while most PDs, including those studied by Turnbull and colleagues, are feedback inhibited by L-Tyr, alignment of the dehydrogenases from the archaeal symbionts predict that they may be insensitive to the end product. Kinetic studies on these archaeal TyrA proteins would provide further insight into the reactions catalyzed by this protein family. Studies by Turnbull and colleagues on *A. aeolicus* TyrA have also shown that this bacterial protein is thermally stable and resistant to proteolytic degradation, a feature that can be explored in TyrA proteins from other thermophilic organisms.

In this chapter we report the first kinetic characterization of TyrA proteins from hyperthermophilic archaea *N. equitans and I. hospitalis.* Predictions based on the amino acid sequence alignment are assessed by functional studies and confirm a preference for $NADP^+$ in the PD reaction as well as insensitivity to the end product L-Tyr. The kinetic parameters for each of the reactions were determined at a salt concentration, pH and temperature that we assessed as near optimal for the PD enzyme. We also show that *N. equitans* PDT is highly sensitive to regulation by the end product L-Phe, and have fit the inhibition data to a model that best

describes L-Phe's interaction with the enzyme. The thermal stability of both archaeal proteins is evaluated by circular dichroism spectroscopy and activity measurments, and is shown to be extraordinary. We also show that IhTyrA is resistant to proteolysis and can be degraded to yield a stable PD domain.

3.1 Experimental procedures

3.1.1 Materials

Prephenate and chorismate were obtained as outlined in 2.1.1. L-tyrosine L-phenylalanine and *m*-fluoro-D, L-tyrosine were obtained from Sigma. Crude L-arogenate (purity ~ 13%) used in specific activity measurements was a kind gift from Carol Bonner. Its concentration was determined enzymatically using the $NADP^+$-specific AD from *Synechocystis* sp. PCC 6803. *E. coli* NAD^+-dependent CM-PD was used to check for the presence of prephenate in the arogenate sample. NAD^+ and $NADP^+$ (free acid) were obtained from Roche. Concentrations of stock substrate solutions were determined using published extinction coefficients [(130)] and/or enzymatic end-point analysis. All other chemical reagents were obtained commercially and were of the highest quality available. Most experiments were performed on the His-tagged NeTyrA or IhTyrA preparation unless stated otherwise. Recombinant *E. coli* CM-PD was produced as described previously [(58)].

3.1.2 Determination of enzyme activity

CM, PD and PDT activities were determined spectrophotometrically using procedures outlined in 2.1.13. Standard reaction conditions were at pH 8.0 and 80°C in a reaction buffer of 50 mM HEPES with 200 mM NaCl unless otherwise indicated. *E. coli* CM-PD activity was assayed as described previously [(26)]. Steady-state kinetic parameters were obtained by fitting initial

velocity data to the Michaelis-Menten equation and other appropriate rate equations using Grafit 7.0 Software (Erathicus) or GraphPad Prism (version 5.01).

To investigate the effect of temperature on TyrA enzyme stability, the enzyme was incubated at 95°C in capped tubes (30 µL enzyme sample per tube) in storage buffer (see 2.1.7). At different time intervals during the incubation, samples were removed, centrifuged for 5 min at 4°C and residual dehydrogenase activity was determined at 80°C with 1 mM prephenate and 1 mM $NADP^+$. Protein concentration was determined after centrifugation to calculate specific activities. Protein concentration was determined as described in 2.1.12.

3.1.3 Effect of temperature on PD activity

To determine the effect of temperature on PD activity, assays were performed in the reaction mixture using TyrA (1 µg/mL of NeTyrA, or 3 µg/mL of IhTyrA) with 0.25 mM prephenate and 2 mM NAD^+, and reaction rates were recorded from 30°C to 95°C. From the data, Arrhenius plots were obtained; activation energy (E_a) values were calculated from the slope using the following equation:

$$K = A * \exp^{(-Ea/RT)}$$

where k is the rate constant, E_a is the activation energy ($kJ.mol^{-1}$), R is the universal gas constant (8.314×10^{-3} kJ mol^{-1} K^{-1}) and T is the temperature in degrees Kelvin. The reactions were also performed in the absence of enzyme and the non-enzymatic rates were subtracted from initial rates of all reactions.

3.1.4 Effects of pH and NaCl on PD activity

The effect of NaCl on the PD activity was monitored using solutions of 50 mM HEPES buffer containing increasing concentrations of salt (0-1 M). The pH of each buffer was checked and adjusted to pH 7.5 at ambient temperature with NaOH as necessary. Reactions were performed at 80°C as described in 2.1.13, but with both substrates added at fixed concentrations of 2 mM NAD^+ and 0.25 mM prephenate. The reaction was initiated with enzyme.

The pH optima for *N. equitans* and *I. hospitalis* PD activities were determined at 80°C by measuring specific activity between pH 5 and 10.5. Assays were performed in a buffer system that contained each of 25 mM MES (pH 5-6.7), Bis-Tris (pH 5.8-7.2), Tris (pH 7-9) and CAPSO (pH 8.9-10.5) supplemented with 200 mM NaCl. PD activity was measured with 0.25 mM prephenate and 2 mM NAD^+ at 80°C as described in 2.1.13

3.1.5 Effect of L-Tyr and L-Phe on TyrA enzyme activity

The effect of L-Phe, L-Tyr or *m*-fluoro-D, L-tyrosine on the CM, PD and PDT activities of NeTyrA and the CM and PD activities of IhTyrA was measured at 80˚C.

Stock solutions of 2.4 mM L-Tyr and 10 mM L-Phe were prepared in standard assay buffer (50 mM HEPES, 200 mM) and the pHs were adjusted to pH 8 at ambient temperature with the careful addition of HCl or NaOH, if required. Assays were performed with different concentrations of L-Tyr and L-Phe by adding the appropriate volume to the reaction buffer to achieve the desired concentration of ligand in solution. The mix was then incubated at 80˚C for 2 min, fixed concentrations of the appropriate substrate(s) were added and after 15 s, the reaction was initiated by the addition of enzyme. Reaction rates were calculated as described above. The effect of L-Tyr (or *m*-fluoro-D, L-Tyr) on CM activity was measured at 290 nm (due to the high

absorbance of L-Tyr at 274 nm) in the presence of either NAD^+ or $NADP^+$. NAD^+ was reported to help L-Tyr bind to *E. coli* TyrA [60]. Assays in the presence of *m*-fluoro-D, L-Tyr were performed as described above. A stock solution of *m*-fluoro-D, L-Tyr in 0.5 M HCl was prepared and stored according to the manufacture's instructions. All stock solutions were stored at 4˚C covered with aluminium foil. When required, a less concentrated stock solution of *m*-fluoro-D, L-Tyr (10 mM) was freshly prepared in assay buffer by dilution and the pH was adjusted to pH 8.

The effect of L-Phe on the velocity of the reaction was obtained by varying prephenate or chorismate. Assays were conducted as described above using concentrations of prephenate, chorismate and L-Phe listed in the Results. Data were plotted using GraphPad Prism (version 5.01) in double reciprocal form and then fitted to the equation describing linear mixed inhibition[157]:

$$v/V_{max} = S/[K_s(1 + I/K_i) + S(1 + I/\alpha K_i)]$$

where v is the initial velocity of the reaction, Vmax is the maximum velocity, [S] represents the concentration of prephenate, K_s is the dissociation constant for prephenate from enzyme (E), [I] is the L-Phe concentration, K_i is the dissociation constant for L-Phe, and the αK_i is the dissociation constant of I from the ESI complex.

E. coli CM-PD activity in the presence of L-Tyr and *m*-fluoro-D, L-Tyr was assayed as described previously[26, 60].

3.1.6 Far-UV circular dichroism spectroscopy

Far-UV circular dichroism (CD) studies were performed on TyrA proteins using a Jasco-815 CD spectropolarimeter equipped with a thermostatically regulated cuvette holder. Protein

samples (0.1–0.3 mg/mL) were exchanged into 50 mM potassium phosphate buffer (pH 7.4) and far-UV CD spectra were recorded from 200-260 nm at 25°C in a 0.2 cm path length cell by averaging 5 wavelength scans from 200-260 nm (1 nm bandwidth) using a response time of 0.25 s, a data pitch of 0.2 nm, a sensitivity of 100 mdeg and a scanning speed of 20 nm/min[52]. All spectra were corrected for contributions by the buffer and smoothed using JASCO software set at the default settings. For variable temperature experiments, changes in ellipticity at 222 nm (1 nm band width) were recorded from 25°C to 95°C with a ΔT of 20°C/h. Spectral scans (200-260 nm) were also recorded at the end of the temperature ramping experiment as described above.

3.1.7 Limited proteolysis with trypsin

Limited proteolysis was conducted on IhTyrA (His tag removed) and *E. coli* CM-PD using trypsin (Roche, sequencing grade, 1.6 µg/8 uL aliquots stored at -80°C). The TyrA protein (1 mg/ml and 0.2 mg/mL, respectively, previously buffer exchanged into 50 mM ammonium bicarbonate pH 8) was incubated with trypsin at a ratio of 50:1 or 250:1 (w/w) at 37°C in the ammonium bicarbonate buffer. At different time intervals during the incubation, the reaction was stopped by adding SDS-PAGE loading buffer, and then boiled for 10 min. The partially digested protein was resolved by 12% SDS-PAGE and stained by Coomassie Blue. The bands representing protein product were excised from the gel, and each band was destained, reduced with dithiothretiol, alkylated with iodoacetomide, digested with trypsin and the peptides were extracted from a gel and analyzed by LC-MS/MS as described in 2.1.14. The partially digested protein was also analyzed by ESI-MS to detect the mass of the stable domains. In that case, at different time intervals during the incubation of TyrA protein with trypsin, sample was collected and immediately subjected to HPLC-MS analysis.

3.2 Results

3.2.1 Effect of pH and NaCl on PD activity

The effects of pH, salt and temperature on the PD activities of both archaeal TyrA enzymes were examined.

The effect of NaCl on the PD activity was monitored using reaction buffer solutions containing increasing concentrations of salt (0-1 M). Reactions were performed at 80°C as described in 2.1.13. The reaction was initiated with either NeTyrA or IhTyrA. IhTyrA demonstrated high activity within a wide range of salt concentrations while NeTyrA was most active at 200 mM NaCl (Fig 3.1A). This salt concentration is similar to the optimal salt concentration determined for the growth of the archaeal organism *N. equitans* (~ 240 mM NaCl)[120]. In the absence of salt, NeTyrA exhibited 57% activity, while the activity of IhTyrA remained close to its maximum. At 1 M NaCl, a decrease in activity was observed for both archaeal TyrAs as NeTyrA and IhTyrA exhibited 46% and 66% of maximal activity, respectively. The profile is comparable to that obtained for *A. aeolicus* PD, which exhibited a maximum activity between 100-250 mM NaCl[52].

The pH optima for *N. equitans and I. hospitalis* PD activities were determined at 80°C by measuring PD activity as described in 3.1.4. The optimal pH for IhTyrA was 8, which is similar to the pH optimum obtained for other PD enzymes such as the monofunctional PD from *B. subtilis* and *A. aeolicus* as well as the bifunctional CM-PD from *E. coli*[52, 100, 158]. The PD activity of NeTyrA, however, increased with pH when tested between pH 5 and 10.5 (Fig 3.1B). This was unexpected and could be related to heterogeneity of the protein sample and/or the effects of the buffer.

Nevertheless, all kinetic assays for both enzymes were performed at pH 8 and 200 mM NaCl. These pH values are much higher than pH optimum for growth of *N. equitans* and *I. hospitalis* (pH 5.5).

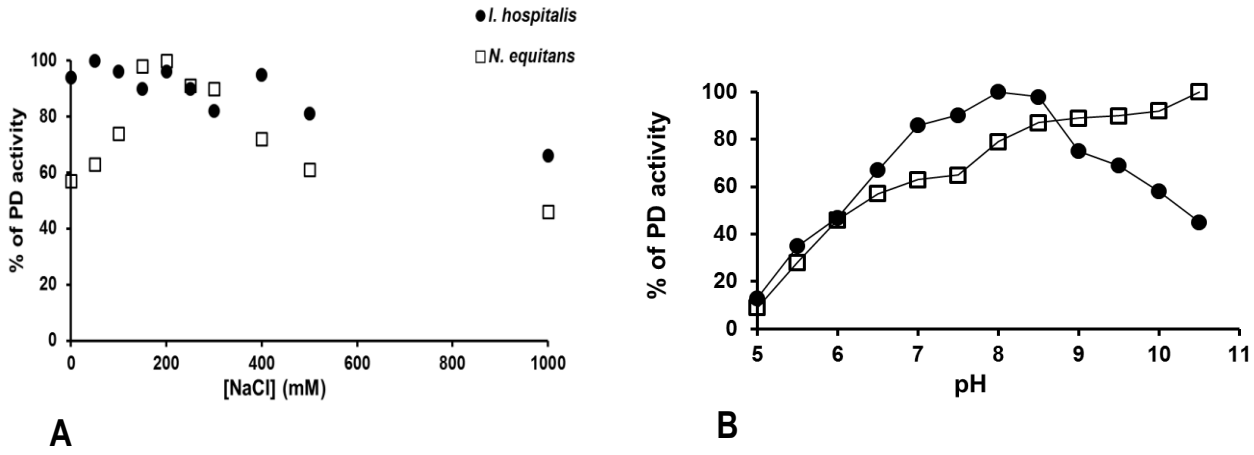


Figure 3.1: The effects of (A) salt and (B) pH on PD activity. PD activity of IhTyrA and NeTyrA was measured with 0.25 mM prephenate and 2 mM NAD^+ at 80°C as described in 3.1.4. For pH dependency studies, PD activity was measured in a buffer system that contained 25 mM MES, Bis-Tris, Tris, CAPSO supplemented with 200 mM NaCl.

3.2.2 Effect of temperature on PD activity

The PD activity of both IhTyrA and NeTyrA PD was low at temperatures below 40°C , but exponentially increased reaching a maximum at 90°C, which is the physiological optimum growth temperature of the symbiotic archaea [(120)]. At 95°C the NeTyrA activity decreased by 20% while IhTyrA activity remained near its maximum. Arrhenius plots were linear from 40° to 85°C for *N. equitans* PD and up to 90°C for *I. hospitalis* PD (Fig 3.2, inset) yielding approximate activation energies of 43 kJ/mol and 46 kJ/mol, respectively. These values are comparable to those obtained for other hyperthermophilic proteins, including the trifunctional TyrA from *A. fulgidus* [(52, 55)].

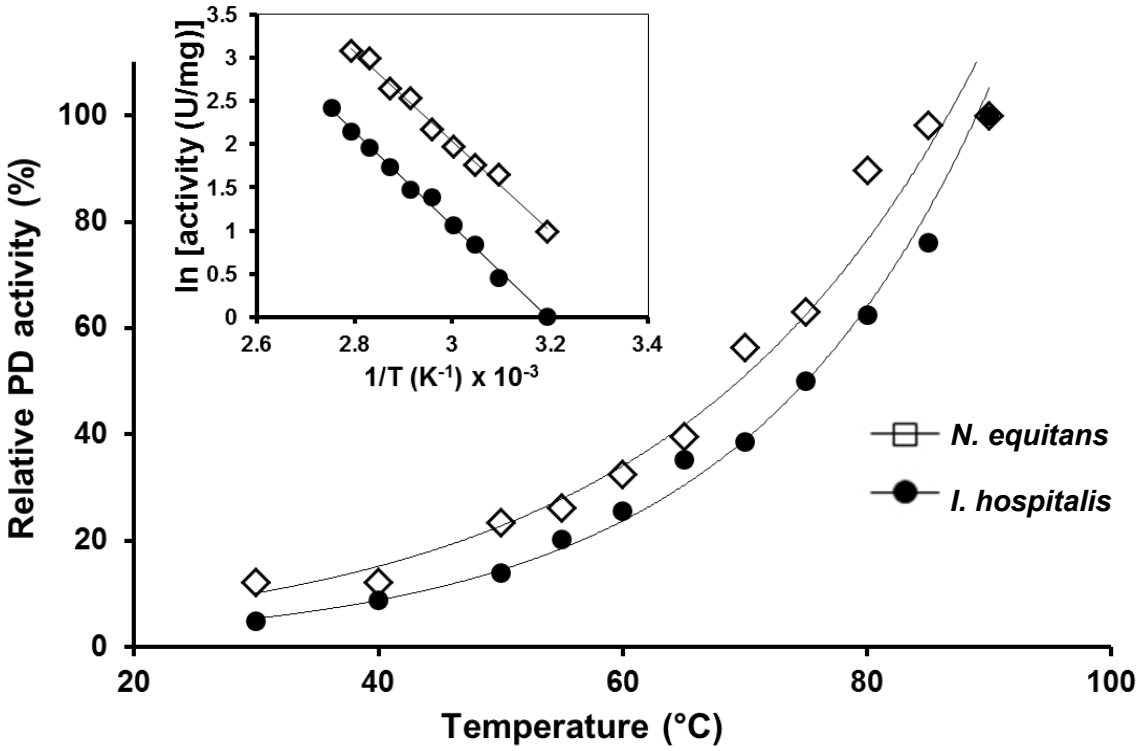


Figure 3.2: Effect of temperature on PD activity. PD activity of NeTyrA and IhTyrA was measured with 0.25 mM prephenate and 2 mM NAD^+ as described in 2.1.13. Activity at 90°C corresponded to 22 U/mg and 16 U/mg for NeTyrA and IhTyrA, respectively. The Arrhenius plot (inset) was generated from measurements over the range of 40 to 85°C for NeTyrA and 40 to 90°C for IhTyrA.

3.2.3 Determination of kinetic parameters for the reactions catalyzed by NeTyrA and IhTyrA

Enzyme assays of the archaeal TyrA proteins confirmed the presence of all predicted catalytic domains: CM and PD activities were detected in the case of IhTyrA while CM, PD and PDT activities were detected with NeTyrA. All enzymatic activities followed Michaelis-Menten kinetics and did not exhibit substrate inhibition. The kinetic parameters obtained from substrate saturation curves of enzymatic reactions performed at 80°C are summarized in Table 3.1. Little or no activity was detected with assays conducted at 30°C. In the case of the PD reaction the fixed substrate in each reaction was maintained at or near a saturating level while varying the

concentration of the other substrate. No significant difference in kinetic parameters was observed between His tagged and tag cleaved IhTyrA (Table 3.1) implying that the N-terminal affinity tag has no effect on this enzyme's function. In the case of NeTyrA, values obtained for the K_M of the substrates with enzyme expressed from the synthesized gene agree with those obtained previously with NeTyrA from the initially cloned gene (Appendix 3A). However, the specific activity (V_{app}) for all the reactions was at least fourfold higher implying a greater TyrA content in the protein obtained from the synthesized gene.

Amino acid sequence alignment with a number of TyrA proteins (Figure 1.8) predicted that both archaeal enzymes would utilize $NADP^+$ in the PD-catalyzed reaction. Both TyrA PD proteins, however, were very active in the presence of either $NADP^+$ or NAD^+ although the relative catalytic efficiency ($V/K_{M\ app}$) for cofactor was approximately fivefold higher with $NADP^+$ (Fig 3.3 and Table 3.1). This increase in reaction rate was due to a tenfold lower apparent K_M for $NADP^+$ versus NAD^+ (Table 3.1). The apparent K_M for prephenate was similar for the two PD enzymes. Preliminary studies indicated that NeTyrA did not utilize L-arogenate with either NAD^+ or $NADP^+$ as a cosubstrate while IhTyrA could use L-arogenate with $NADP^+$ but very poorly (specific activity with L-arogenate was $\sim$ 200-fold lower than with prephenate). The broad cofactor specificity is very unusual for prephenate dehydrogenases that generally use NAD^+. CM and PDT ($V/K_{M\ app}$) activities calculated for both IhTyrA and NeTyrA were within tenfold of the PD activity with $NADP^+$ as the cofactor, although CM activity from IhTyrA was fivefold higher than for NeTyrA (Table 3.1).

	Mutase activity			Dehydrogenase activity						Dehydratase activity		
Variable S	Chorismate			Prephenate			$NADP^+$ (NAD^+)			Prephenate		
Fixed S	------			$NADP^+$ (NAD^+)			Prephenate			-------		
Protein	K_M (μM)	V_{app} (U/mg)	$V/K_{M\,app}$ ($M^{-1}U/mg$)	K_M (μM)	V_{app} (U/mg)	$V/K_{M\,app}$ ($M^{-1}U/mg$)	K_M (μM)	V_{app} (U/mg)	$V/K_{M\,app}$ ($M^{-1}U/mg$)	K_M (μM)	V_{app} (U/mg)	$V/K_{M\,app}$ ($M^{-1}U/mg$)
His tagged PD-CM-PDT	1023 ± 25	21 ± 2	2.1×10^4	730 ± 50 (690 ± 100)	67 ± 1 (125 ± 1)	9.1×10^4 (1.8×10^5)	360 ± 30 (3300 ± 600)	80 ± 2 (148 ± 2)	2.2×10^5 (4.5×10^4)	205 ± 24	30 ± 0.9	1.5×10^5
His tagged CM-PD	435 ± 34	47 ± 4	1.1×10^5	443 ± 105 (319 ± 35)	23 ± 3 (21 ± 2)	5.1×10^4 (6.5×10^4)	62 ± 15 (561 ± 27)	21 ± 3 (30 ± 2)	3.4×10^5 (5.3×10^4)	ND	ND	ND
Untagged CM-PD	386 ± 36	71 ± 4	1.8×10^5	426 ± 49	24 ± 1	5.6×10^4	ND	ND	ND	ND	ND	ND

Table 3.1: Steady-state kinetic parameters for the reactions catalyzed by NeTyrA and IhTyrA. Assays were performed as indicated in 2.1.13 and 3.1.2. Values were calculated from the initial rates using at least 6 substrate concentrations. CM reaction - chorismate was varied from 0.025 - 1.5 mM. PD reaction – prephenate was varied from 0.02 - 4 mM (0.025 - 8 mM with PD-CM-PDT) at a fixed concentration of 2 mM NAD^+ (40 mM used with PD-CM-PDT) or 0.5 mM $NADP^+$ (2 mM used with PD-CM-PDT). When varying the cofactor concentration of $NADP^+$ (0.01 - 4 mM) or NAD^+ (0.01 - 10 mM), prephenate was fixed at 2.5 mM. PDT reaction – prephenate was varied from 0.1 - 3.3 mM. Values in parentheses were obtained from assays performed with NAD^+. ND- not determined. Kinetic values (± S.E.) are an average of at least two different experiments each done in duplicate.

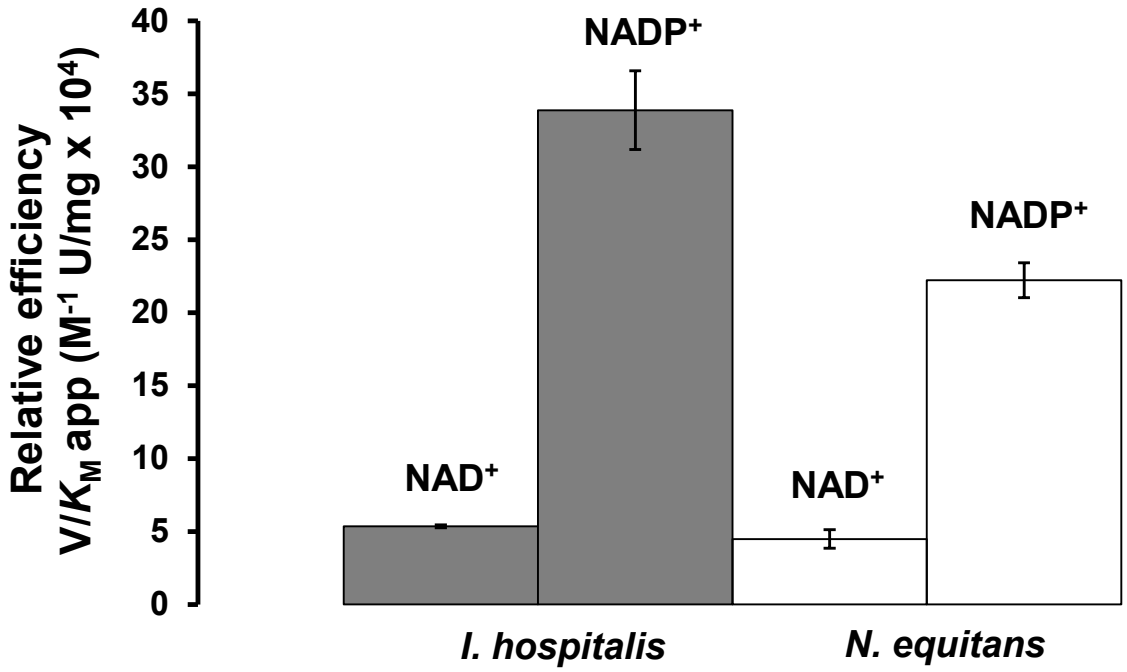


Figure 3.3 Cofactor specificity of the PD reaction catalyzed by IhTyrA and NeTyrA. Values for the kinetic parameters (V/K_M app for cofactor) were obtained as specified in Table 3.1.

3.2.4 Effect of L-Tyr and L-Phe on the activities of NeTyrA and IhTyrA

The enzymatic activities of NeTyrA and IhTyrA were examined in the presence of end products L-Tyr and L-Phe and the results are shown in Figure 3.4. The figure shows that the CM and PD activities of TyrA proteins from both archaeal species were not affected by 1 mM L-Tyr. This insensitivity to L-Tyr is unusual and contrasts most PD proteins characterized to date. Therefore, we also tested a fluorinated L-Tyr analog, *m*-fluoro-D, L-tyrosine (DL-FTyr) which is reported to be a more potent inhibitor of PD activity than L-Tyr [(159)]. Nevertheless, no inhibition was observed of CM or PD activities in the presence of 1 mM DL-FTyr (Figure 3.4). The lack of feedback inhibition is contrary to tyrosine-sensitive *E. coli* TyrA that was markedly inhibited both by L-Tyr and its analog (Figure3.4). The CM and PD activities were also not affected by 1 mM L-Phe.

While L-Tyr and DL-FTyr did not influence *N. equitans* PDT activity at concentrations up to 1 mM, PDT was markedly inhibited by L-Phe (Fig 3.4). PDT lost ~ 90% of its activity in the presence of 5 µM L-Phe. L-Phe inhibiton was further examined with prephenate as a variable substrate. Double reciprocal plots for all reactions in the presence of L-Phe were linear (Fig 3.5). Thus, PDT showed no cooperative effects in its interactions with prephenate in the presence of L-Phe under our experimental conditions. Initial velocity data were fit to the equations describing competitive inhibition, uncompetitive and non-competitive inhibition. However, the best fit was obtained to a model describing a mixed type of non-competitive inhibition. This is consistent with the formation of an enzyme-prephenate-L-Phe complex in addition to an enzyme-prephenate complex. The model also predicts an αK_i term (an interaction term). The PDT- L-Phe complex displayed a fivefold lower affinity for prephenate than did the free PDT while the PDT-prephenate complex showed a fivefold lower affinity for L-Phe than the free enzyme (K_i = 0.8 µM, αK_i = 4.4 µM).

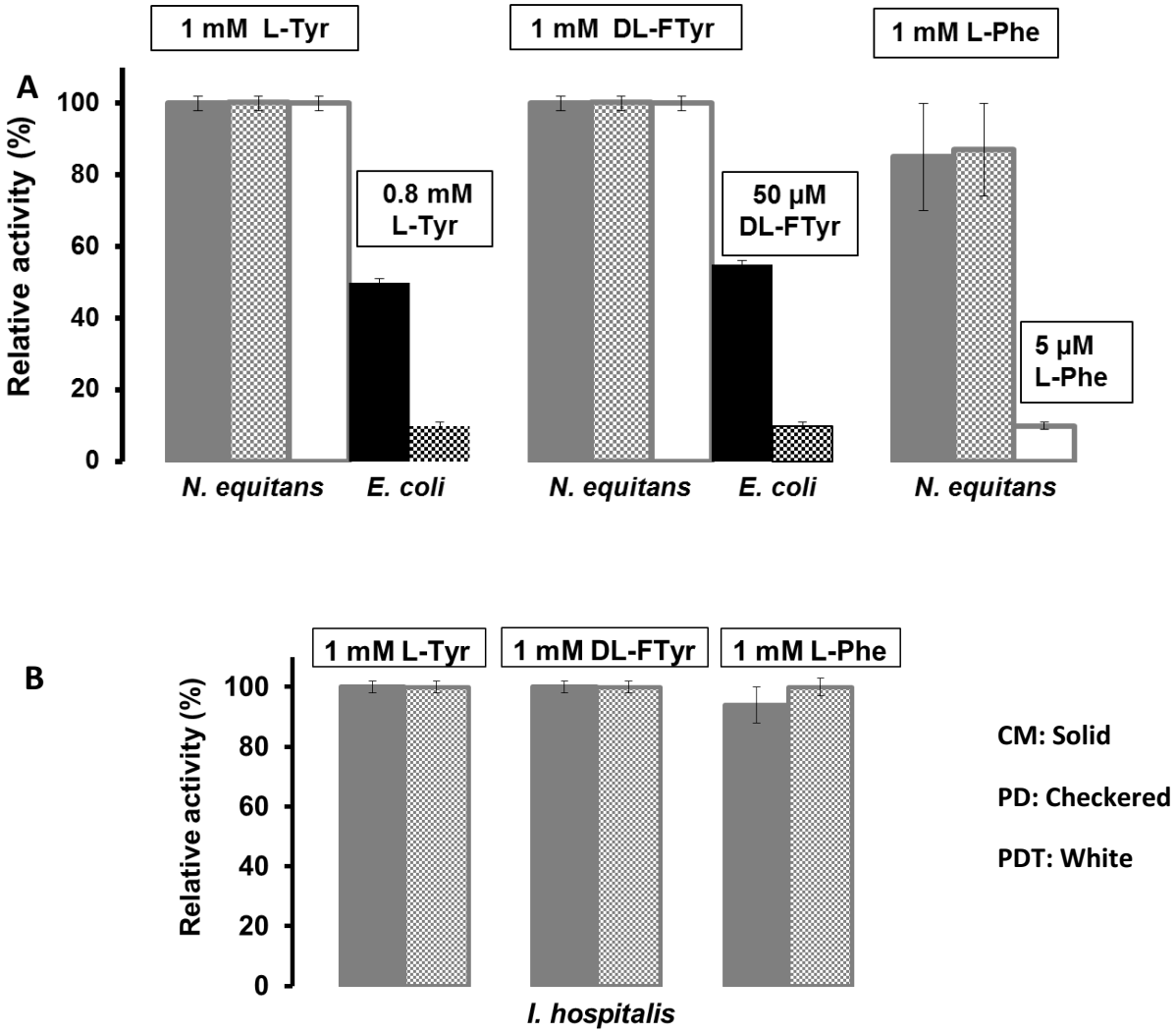


Figure 3.4: Profile of the inhibition of TyrA enzymes by end products and an end product analog. Enzyme activities were determined for (A) *N. equitans* PD-CM-PDT (gray) and *E. coli* CM-PD (black) and (B) *I. hospitalis* CM-PD. For the chorismate mutase reaction 1 mM chorismate was used in the presence or absence of cofactor (2 mM NAD^+ or $NADP^+$ for *N. equitans* enzyme). For the prephenate dehydrogenase reaction, prephenate (0.5 mM) was used with either 10 mM NAD^+ or 2 mM $NADP^+$. For the PDT reaction 0.5 mM prephenate was used. *E. coli* CM-PD served as a control to represent sensitivity to L-tyrosine. *E. coli* CM-PD was assayed as described in Turnbull and Morrison [(60)], except that fluoro tyrosine (DL-FTyr) was used, while data for the *E. coli* CM-PD inhibition by L-Tyr were obtained previously[(60)]. Error bars represent standard error of the mean of at least two separate determinations.

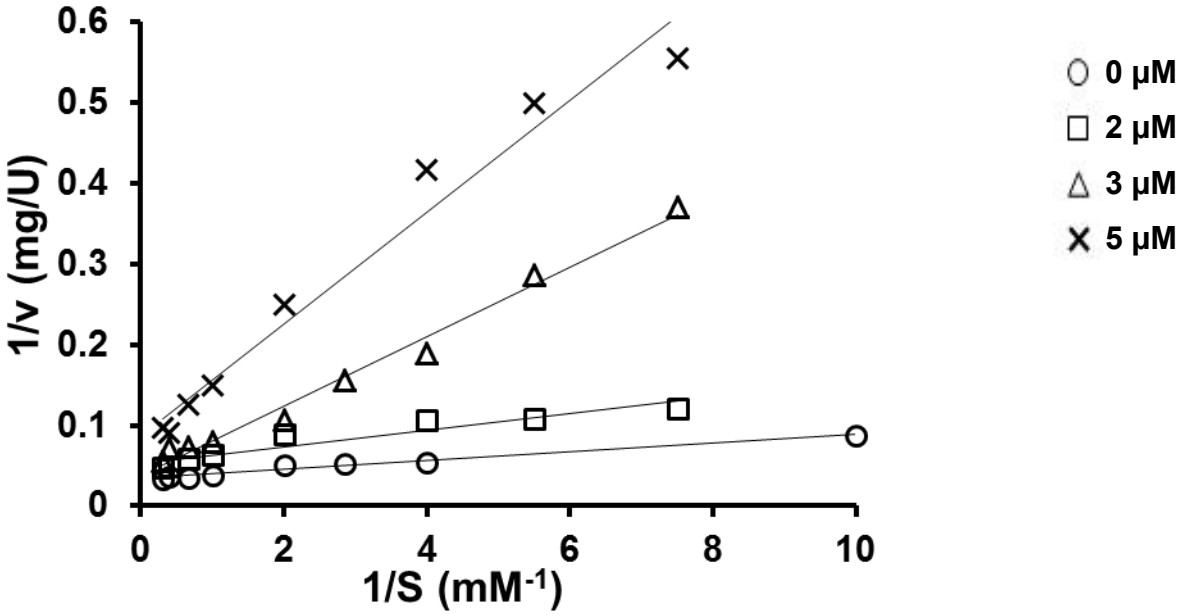


Figure 3.5: Double reciprocal plots for the inhibition of PDT activity of NeTyrA by L-Phe. Data were analyzed using curve fitting options provided by GraphPad Prism (V5.01). Initial velocity data varying prephenate in the presence of L-Phe were fit to the equation described by mixed-type inhibition. The following values were obtained from the analysis: K_i = 0.8 ± 0.2 µM, K_s = 193 ± 36 µM, V_{max} = 31 ± 1.3 U/mg and α = 5.5.

3.2.4 Thermal stability of TyrA proteins

The hyperthermophilic archaea *N. equitans* and *I. hospitalis* thrive at high temperatures and therefore are predicted to possess thermally stable proteins including the TyrA enzyme. Accordingly, variable temperature far UV-CD spectroscopy was used to investigate the thermostability of the archaeal TyrA proteins. The far-UV CD spectra of untagged NeTyrA (thrombin treated, reapplied on Ni-NTA) and His-tagged IhTyrA recorded at 25°C showed two minima at 208 nm and 222 nm indicating that the proteins within these preparations possessed a significant content of α-helical structure (Fig 3.6). These data agree with helical contents calculated from the primary sequence of the archaeal TyrAs using secondary structure prediction

programs (SCRATCH (http://www.igb.uci.edu/servers/psss.html) [160] and Phyre 2 (www.sbg.bio.ic.ac.uk/phyre2/) [161] that were between 50% and 60%. The unfolding of TyrA proteins by thermal denaturation was followed by measuring changes in ellipticity at 222 nm when samples were gradually heated from 25°C to 95°C (Figure 3.6). Little change was observed and suggested that the midpoint of the unfolding (T_m) occurs at temperatures higher than 95°C. Additionally, scans performed at the end of the variable temperature experiment showed that the proteins had retained their helical character, although the change in ellipticity for the NeTyrA sample was consistent with the temperature-induced denaturation of some *E. coli* proteins. These results are as expected for highly thermostable proteins. By comparison, the unfolding of mesophilic *E. coli* CM-PD generated a sigmoidal curve with a well-defined T_m value of 63°C.

In some cases, hyperthermophilic proteins can be resistant to chemical denaturants such as guanidine hydrochloride (Gdn-HCl). To test if IhTyrA is stable to chemical denaturation we performed preliminary studies which showed that when assayed in the presence of 3 M Gdn-HCl at ambient temperature, the enzyme retained full dehydrogenase activity after incubation for 30 min in the reaction buffer and retained 20% of activity after 24 h of incubation (data not shown).

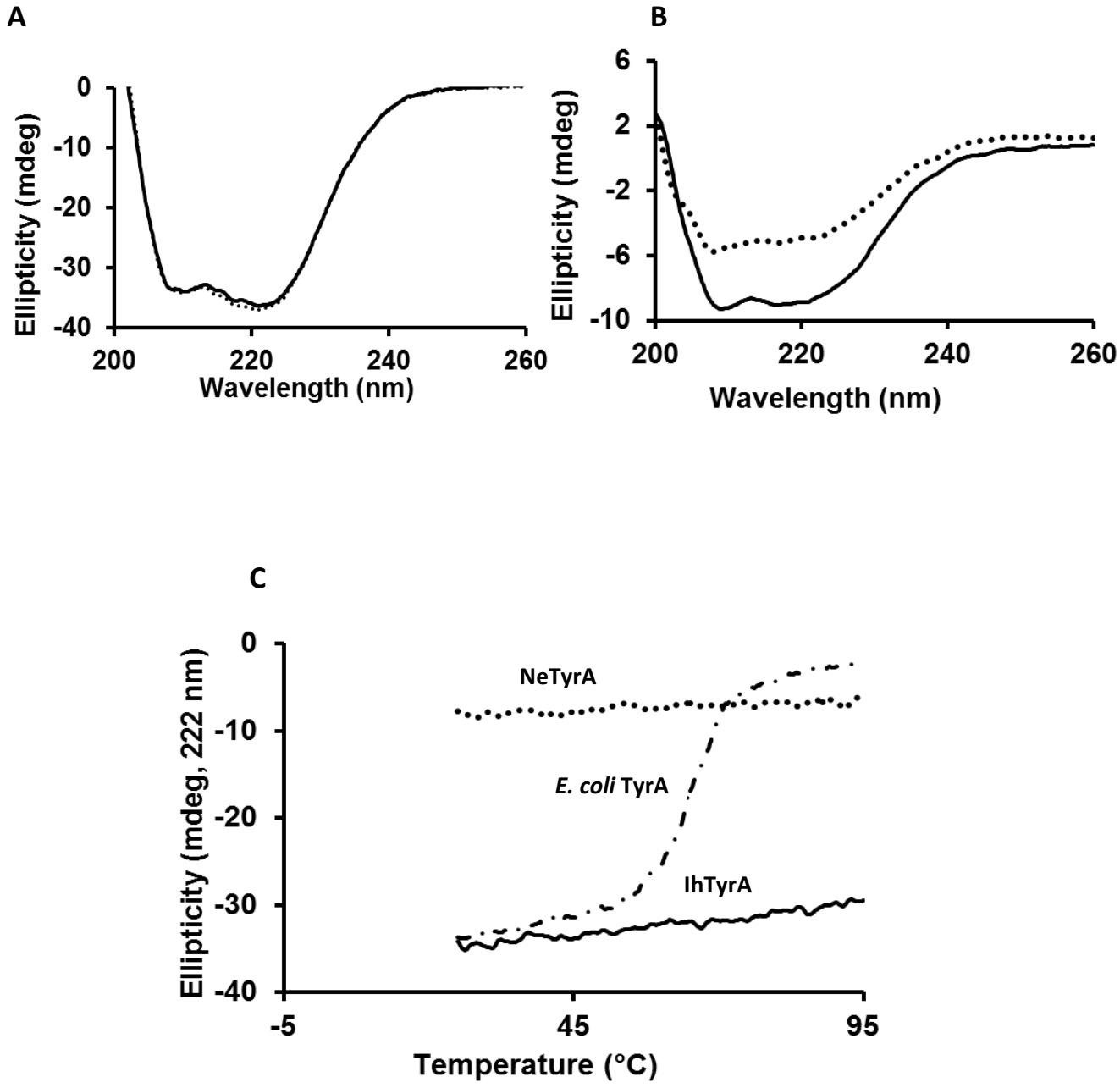


Figure 3.6: Far-UV CD spectra of TyrA proteins. CD spectra of protein samples (0.1-0.3 mg/mL) containing IhTyrA (A) and NeTyrA (B) in 50 mM potassium phosphate buffer (pH 7.4) were recorded at 25°C (solid line and after thermal unfolding at 95°C (dotted line). In (C) thermal denaturation experiments with IhTyrA (solid line) NeTyrA (dotted line) and *E. coli* CM-PD (dashed) were carried out by following the ellipticity at 222 nm from 25-95°C. *E. coli* CM-PD was used as a positive control for thermal unfolding.

To investigate further the thermal stability of the TyrA proteins, the enzymes were incubated at 95°C while monitoring dehydrogenase activity at 80°C and the soluble protein content at various time intervals (Fig 3.7). The same samples were also analyzed by SDS-PAGE (Fig 3.8 A and B). The specific activity of *N. equitans* PD increased twofold after 1 h of incubation while denaturing gel electrophoresis revealed that the amount of soluble protein at the molecular weight of NeTyrA (~ 71 kDa) diminished significantly (Fig 3.8 A). These data suggest that much of the protein in the sample at that size was *E. coli* contaminants, which precipitated at high temperature. Interestingly, the activity of *I. hospitalis* PD also increased about twofold during the first 1-3 h of incubation (Fig 3.7), but without a significant change in the amount of soluble protein. This is also reflected in the denaturing gel (Fig 3.8 B), which shows most protein bands are retained even after 50 h of incubation.

Both NeTyrA and IhTyrA appeared to be extraordinary thermostable retaining full dehydrogenase activity after 50 h incubation at 80°C (data not shown) and greater than 50% of PD specific activity after 50 h at 95°C (Fig 3.7). IhTyrA appeared more thermally stable than NeTyrA, however, and retained 92% of initial PD activity after 50 h at 95°C.

Additionally, SDS-PAGE analysis of NeTyrA revealed the presence of another thermally stable species at ~ 20 kDa whose amount gradually decreased with incubation time at 95°C. In order to identify proteins in this fraction, along with protein migrating at ~ 71 kDa, tandem MS of tryptic digestions of the appropriate gel band (Fig 3.8 A) was performed (Appendix 3B). Analysis of a 3 h heat-treated sample identified chaperone SlyD (predicted mass-24 kDa) with the highest confidence. A portion of the PD domain of NeTyrA in the ~ 20 kDa band was also detected. In the ~ 71 kDa band peptides from NeTyrA were identified with the highest confidence but tandem MS also detected a significant presence of chaperone DnaK. These data

reveal that prolonged heating greatly assists in the purification of *N. equitans* TyrA from *E. coli* proteins but is not effective in eliminating heat-stable chaperones DnaK and SlyD.

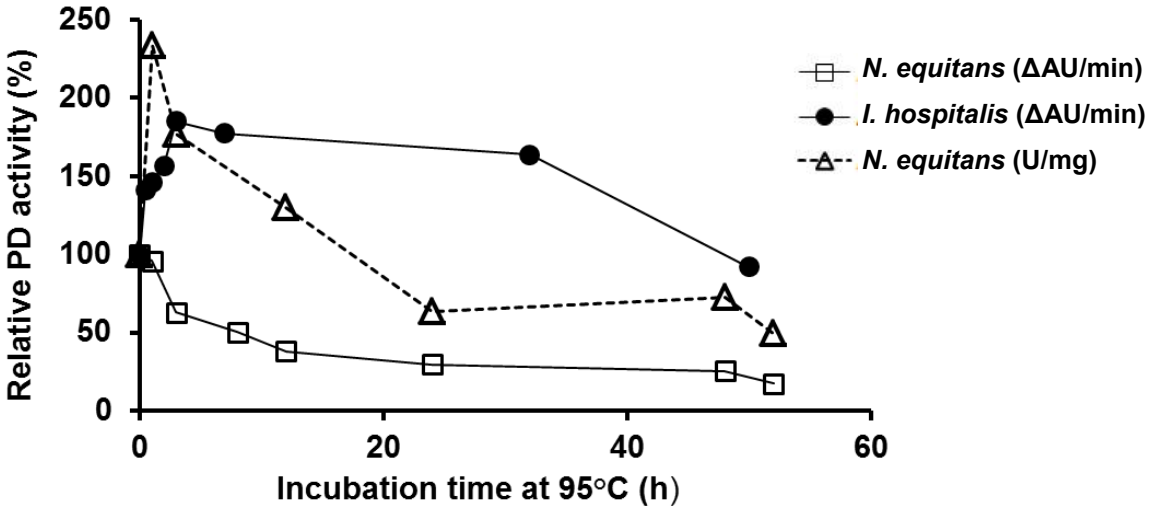


Figure 3.7: Thermal stability profile of NeTyrA and IhTyA. Residual prephenate dehydrogenase activity was measured at 80°C with 1 mM prephenate and 1 mM $NADP^+$ after incubation at 95°C for increased length of time. Values of S.A. for IhTyrA are not shown although they follow the same pattern as ΔAU/min.

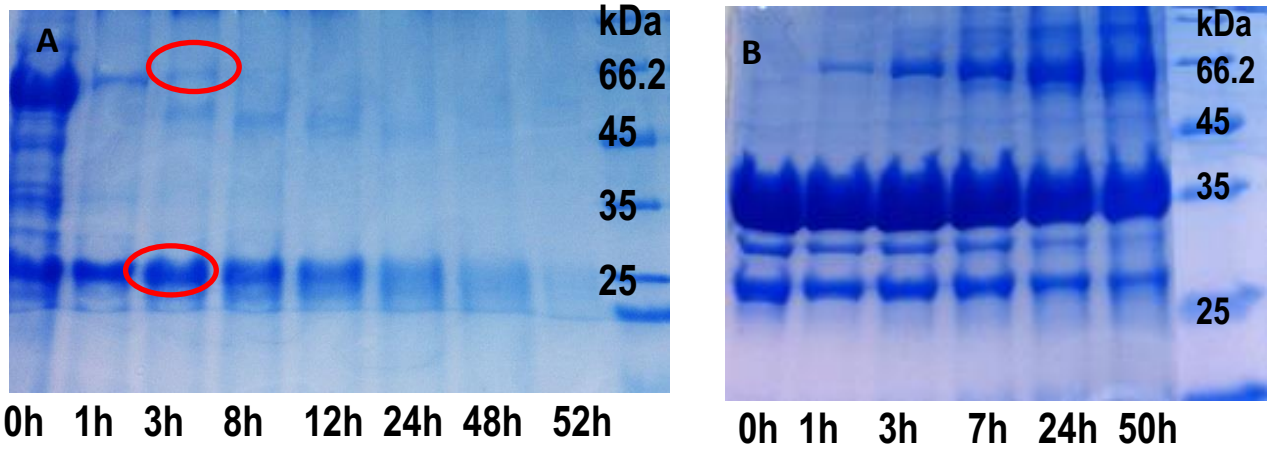


Figure 3.8: SDS-PAGE analysis of heat-treated NeTyrA and IhTyrA. Samples (1 mL) of NeTyrA (A) and IhTyrA (B**)** were incubated at 95°C at various times, centrifuged and the soluble fractions were analyzed by SDS-PAGE and PD activity measurements. For the analysis of NeTyrA, circled regions in the gel were subjected to in-gel tryptic digestion and tandem MS analysis.

3.2.5 Proteolytic susceptibility of IhTyrA protein

A time-dependent limited proteolysis experiment was conducted to assess the sensitivity of IhTyrA to proteolytic digestion and to identify protease-resistant fragments. Protein was incubated in the presence of trypsin from 0-96 h and the samples were analyzed by SDS-PAGE. As shown in Fig 3.9, *I. hospitalis* CM-PD was very resistant to proteolysis by trypsin. In contrast to *E. coli* CM-PD, which was completely digested by 24 h (Lanes 7 and 8), IhTyrA retained most of its intact structure after 24 h in the presence of trypsin (Lane 3), even when the amount of trypsin was increased fivefold (Lanes 3 and 6). Additionally, formation of a stable domain at ~ 30 kDa size was noted. ESI-MS analysis of samples subjected to limited proteolysis established the domain's mass of 29 613 Da and showed that its relative abundance increased with incubation time (Fig 3.10). From the the exact mass and *in-silico* tryptic digestion information of IhTyrA (Appendix 2B), it was calculated that this fragment likely corresponds to a C-terminal PD domain, which is missing the first 88 amino acids of the protein's CM domain. The cleavage site lies between Lys88 and Ile89 in the primary sequence of IhTyrA and is consistent with trypsin's specificity for cleaving at the carboxy end of amino acids Arg and Lys. With time, additional fragments appeared as a result of proteolysis at the protein's C-terminus (Fig 3.10, C). In summary, the highly thermostable IhTyrA seems to be also highly resistant to proteolysis with its PD domain much less sensitive to tryptic digestion than the CM domain.

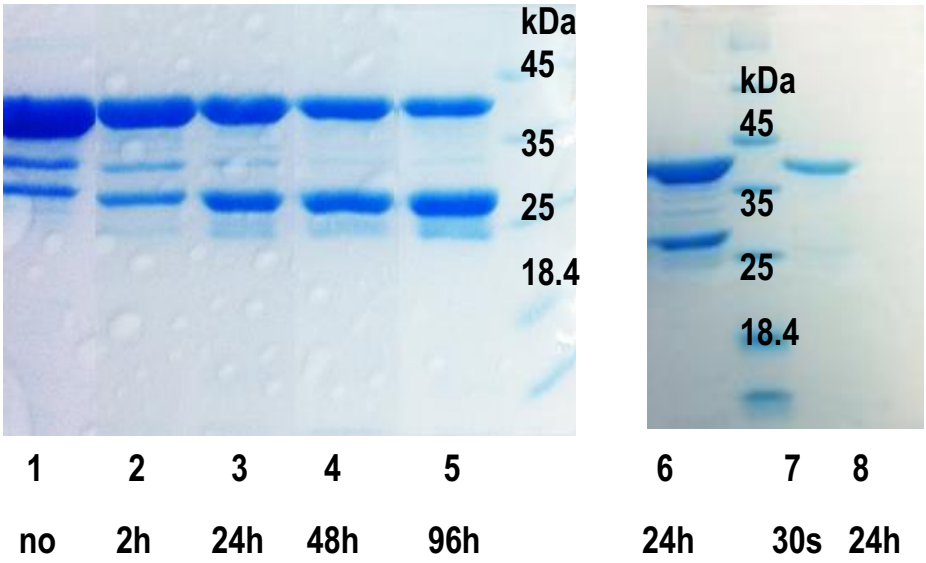


Figure 3.9: SDS-PAGE analysis of IhTyrA and *E. coli* TyrA after limited tryptic digestion. Lanes 1-5 show the digestion pattern of IhTyrA in the absence or presence of 1:50 w/w trypsin to TyrA. Lane 1: in absence of trypsin, Lane 2: after 2 h incubation, Lane 3: after 24 h incubation, Lane 4: after 48 h incubation, Lane 5: after 96 h incubation. Lanes 6-8 show digestion pattern of TyrAs in the presence of 1:250 (w/w) trypsin to TyrA. Lane 6: IhTyrA after 24 h incubation, Lane 7: *E. coli* TyrA after 30 s incubation, Lane 8: *E. coli* TyrA after 24 h incubation.

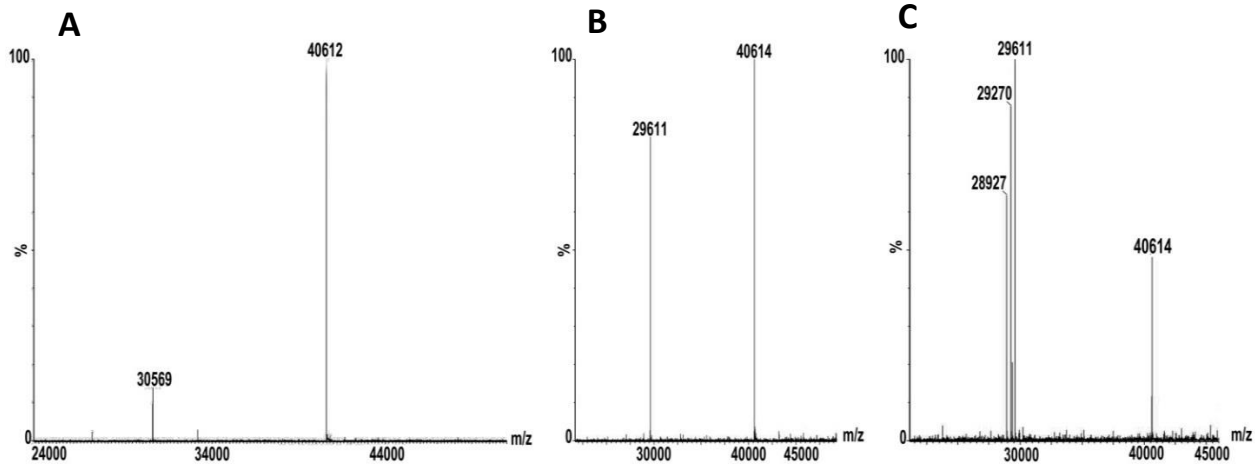


Figure 3.10: **ESI-MS analysis of IhTyrA after limited tryptic digestion.** A-C shows the digestion pattern of IhTyrA in the absence or presence of 1:50 (w/w) trypsin to TyrA (A) in absence of trypsin (B) after 24 h of incubation and (C) after 96 h incubation. Mass of 29 611 is correlated with protein fragment after trypsin cleavage between Lys88 and Ile89, masses of 29 270 and 28 927 are correlated with additional cleavage at position Arg345 and Arg343. The intact protein is shown at mass of 40 612 (40 614).

3.3 Discussion

The archaeal TyrA proteins show high thermal stability

Hyperthermophilic bacteria and archaea thrive at the high temperatures of their exogenous environment. As such, they possess highly thermally stable proteins. The TyrA enzymes from the archaeal symbionts provide a good example of this adaptation.

As deduced by far-UV CD spectroscopy, the melting temperatures (T_m) of NeTyrA and IhTyrA were greater than 95°C (Fig 3.6). This is not surprising since the physiological optimum growth temperature of their symbiotic archaeal *N. equitans* and *I. hospitalis* is near 90°C[120]. Similarly, a high T_m value (>100°C) was reported for monofunctional PD from the hyperthermophilic bacterium *A. aeolicus*[52]. For comparison, mesophilic *E. coli* CM-PD exhibited much lower T_m of 63°C (Fig 3.6). Far-UV CD analysis, which provides valuable information regarding the global secondary structure of proteins, indicated that at 25°C both archaeal TyrA protein samples possessed a high abundance of α-helical structure (Fig 3.6). For NeTyrA, however, the spectrum also represents the contribution of coeluting helical *E. coli* proteins including DnaK chaperone. Nonetheless, the high α-helical content is typical for many CM, PD, and PDT proteins from other mesophilic and hyperthermophilic organisms [75, 89, 92, 107, 112]. Models derived of IhTyrA's CM domain (Appendix7A) and PD domain (Appendix 7B) show that both portions of the protein contribute to its helical structure and support the finding from the far-UV CD measurements.

The thermal stability of NeTyrA and IhTyrA was further explored by measuring time-dependent changes in PD activity during incubation of the enzyme at 95°C. Both NeTyrA and IhTyrA were remarkably thermally stable; the enzymes retained greater than 50% of PD activity after 50 h at 95°C although IhTyrA appeared more thermally stable than NeTyrA, retaining 92% of initial PD activity after 50 h incubation (Fig 3.7). In addition, PD activities increased about twofold

during the first 1 to 3 hours of incubation at high temperature implying at least for IhTyrA that there is a temperature-induced conformational change that is beneficial for the efficient processing of the heat labile substrates. A time-dependent increase in activity was not observed at high temperatures with hyperthermophilic *A. aeolicus* PD implying different mechanisms adopted by the bacterium for substrate processing. The high degree of thermostability observed for the archael TyrAs contrasts the findings for TyrA proteins from other organisms. The PD from *A. aeolicus*, for example, lost 50% of its activity after only 2 h of incubation at 95°C [(52)] while only modest half-lives at 40°C have been found for mesophilic *E. coli* CM-PD[(52)] and for other TyrA proteins[(29, 51, 162)]. In fact, the archaeal enzymes are the most thermally stable TyrA proteins reported to date. Our findings are in agreement with those reported for other proteins isolated from hyperthermophilic archaea such as the alanine dehydrogenase from *A. fulgidus* ($T_{1/2}$ 90°C = 50 h in the presence of 1.5 M KCl [(163)]), protease from *Pyrococcus furiosus* ($T_{1/2}$ 98°C = 33 h in the presence of SDS) [(164)], ferredoxin from *Thermococcus litoralis* ($T_{1/2}$ 95°C > 24 h) or amylopullulanase from *Thermoproteus Tenax* ($T_{1/2}$ 98°C = 20 h) [(165, 166)] .

What is the basis for the differences in thermal stability of archaeal TyrAs from *I. hospitalis* and *N. equitans* and the bacterial enzyme from *A. aeolicus*? The difference in the degree of thermostability could be attributed to the diverse mechanisms evolved in archaeal and bacterial proteins to resist heat [(167, 168)]. Evolutionary studies and analysis of structures and sequences of hyperthermostable archaeal and bacterial proteins revealed that proteins from hyperthermophilic archaea that originated in an extreme environment employ a "structure-based" mechanism for thermostabilization; they are significantly more compact and hydrophobic than their mesophilic homologs [(126)]. In contrast, proteins from hyperthermophilic bacteria, that are thought to have originated in moderate temperatures and later recolonized a hot environment by acquiring genes for thermostability through lateral gene transfer from the

archaea, use a "sequence-based" mechanism [126]. While the bacterial hyperthermophilic proteins do not show significant structural differences from mesophilic homologues, they show a small number of strong interactions, such as ion-pair or H-bonding networks. The lateral gene transfer is evident in *A. aeolicus* since 16% of its genes encode proteins that are more similar to archaeal than to bacterial proteins[168].

Our limited biophysical data to date for the archaeal TyrA proteins only partially support a "structure-based" mechanism. Limited tryptic digestion clearly showed that IhTyrA is very resistant to proteolysis (Fig 3.9) compared to *E. coli* CM-PD, likely through a more compact structure (see discussion below). However the sedimentation velocity experiments yield average molecular weights that are in close agreement with the protein's size based on its primary sequence (Fig 2.12). Additionally, the aliphatic indices[169] of NeTyrA and IhTyrA (112 and 98, respectively) are not markedly different from the values for the mesophilic bifunctional CM-PDs from *E. coli* and *H. influenzae* and from *A. aeolicus* PD (a range 96-103).

Additional factors thought to contribute to such increased thermostability were elucidated through an extensive analysis of proteins from hyperthermophilic and mesophilic organisms. These studies highlighted the increase in proline residues[170, 171] that restrict the conformational freedom of a protein, a number of leucine to isoleucine substitutions, as noted for *M. jannaschii* CM[172], as well as the increase in the proportion of charged versus polar (uncharged) amino acids, especially in the proportion of solvent accessible charged residues at the protein surface as noted in many proteins from hyperthermophilic organisms[125, 170, 171]. Sequence analysis of both archaeal TyrA proteins did not reveal any increase in the proline residues or change in leucine:isoleucine ratio when comparing with the sequence of *A. aeolicus* PD or even with *E. coli* CM-PD. However, a higher proportion of glutamate and lysine residues was observed in the archaeal TyrAs. The proportion of glutamate residues in IhTyrA and NeTyrA, for

example is 11.9% and 9.8%, respectively, which is higher than in *A. aeolicus* PD (8%) and is significantly higher than in *E. coli* CM-PD (3.7%). The crystal structure of *A. aeolicus* Δ19PD complexed with NAD^+ revealed a large ionic network formed by Glu275, Glu278 and Lys285 from one subunit and the same residues from the other subunit, a structural feature which is thought to contribute to the thermostability of the enzyme[89]. Solving the crystal structure of IhTyrA proteins from the archaeal symbionts will provide insights into the structural basis for such an exceptional thermostability of these proteins.

In some cases, factors responsible for the thermal stability of hyperthermophilic proteins also confer resistance to other denaturants such as guanidine hydrochloride (Gdn-HCl), SDS, organic solvents, and to proteolysis [164, 173-175]. For example, the unfolding of ribonuclease from the hyperthermophilic archaeon *Sulfolobus tokadii* in the presence of 7 M of Gdn-HCl reached equilibrium only after 8 h [175]. Preliminary chemical denaturation studies of IhTyrA revealed that in the presence of 3 M Gdn-HCl at ambient temperature the enzyme retained full dehydrogenase activity during the first 30 min and retained 20% of activity after 24 h incubation. Interestingly, this archaeal TyrA is much more stable than the bacterial PD from hyperthermophilic *A. aeolicus* which lost most of its activity in the presence of 3 M Gdn-HCl [52]. A thorough effect of chemical denaturant (s) on the unfolding of IhTyrA is yet to be investigated.

I. hospitalis TyrA shows high resistance to proteolysis

Limited tryptic digestion of purified IhTyrA revealed that this protein is resistant to proteolysis, retaining significant intact mass even after a prolonged incubation with trypsin, in marked contrast to mesophilic *E. coli* CM-PD. This implies that most of the trypsin recognition sites (9 in the CM domain and 22 in the PD domain, Appendix 2B) are not easily accessible as

would be the case in a more tightly folded protein. Interestingly, SDS-PAGE analysis identified a stable ~ 30 kDa fragment from partial digestion with trypsin. Mass spectrometry analysis indicated that the stable fragment belongs to the C-terminal PD domain of the protein, which is missing the first 88 residues. The site cleavage (between Lys88 and Ile89) is likely at a more accessible junction between the CM and PD domains. As no additional fragments (~ 10 kDa) were observed during tryptic digestion, this implies that the remaining CM domain is significantly less stable than the PD portion. Interestingly, limited tryptic proteolysis of the thermostable monofunctional PD from *A. aeolicus* has also yielded a stable C-terminal domain, which was active and yielded diffraction quality crystals[(89)]. Additionally, the only reported crystallizable form of the bifunctional *H. influenzae* CM-PD is one in which the first 80 residues comprising the CM domain are deleted[(92)]. Thus, it is possible that this stable truncated IhTyrA could be active and may be a good candidate for crystallization trials. These questions are addressed in Chapter 4.

Kinetic properties of arhaeal TyrA enzymes

Archaeal TyrA proteins overexpressed and purified as described in Chapter 2, possessed enzymatic activities confirming the presence of all predicted catalytic domains: IhTyrA showed CM and PD activities while CM, PD and PDT activities were detected with NeTyrA. Kinetic analysis provided further evidence that the archaeal TyrA retain functional properties predicted for hyperthermophilic proteins. While little or no activity was detected with assays conducted at 30°C, both IhTyrA and NeTyrA PD activities were at a maximum at 90°C, which is the physiological optimum growth temperature of the symbiotic archaea [(120)]. This behaviour is typical to hyperthermophilic enzymes that normally exhibit much lower activity at ambient temperatures and higher activity at elevated temperatures[(176)]. At 95°C the NeTyrA activity decreased by 20% while IhTyrA activity remained near its maximum. These values are

comparable to those obtained for other hypertherthermophilic TyrA proteins such as the monofunctional PD from *A. aeolicus* and the trifunctional PD-CM-PDT from *A. fulgidus* [52, 55]. The Arrhenius plots for both archaeal TyrAs were linear over the temperature ranges selected under our assay conditions (Fig 3.2, inset), which is similar to those reported for *A. aeolicus* and *A. fulgidus* PDs, indicating a single rate-limiting step in the dehydrogenase reaction for all the four hyperthermophilic enzymes. Studies on *E. coli* CM-PD have established catalysis as a single rate-limiting step in the PD-catalyzed reaction[26] . In contrast, studies on *A. aeolicus* PD revealed a 10-fold difference between the binding constant (K_d) and K_M for NAD^+ suggesting that a step other than catalysis may be rate limiting[52]. Whether catalysis is the rate-limiting step in reactions catalyzed by archaeal *N. equitans* or *I. hospitalis* PDs is yet to be investigated.

The kinetic parameters determined for enzymatic reactions (Table 3.1) showed that both IhTyrA and NeTyrA were comparably efficient catalysts; they exhibited similar efficiency in the mutase and dehydrogenase reactions and the efficiency of NeTyrA dehydratase reaction was within the range of its other activities. Additionally, kinetic parameters for the three enzyme-catalyzed reactions were, for the most part, within the range of those reported for PD, CM and PDT enzymes from other organisms[52, 57, 102]. Interestingly, *N. equitans* which is thought to be a parasite of *I. hospitalis* with regard to other metabolic pathways seems to be an efficient, autonomous producer of L-Phe and L-Tyr. The only other trifunctional TyrA characterized to date is from the archaeon *A. fulgidus* [55], which also displays the same putative domain assembly as NeTyrA and is of a similar monomer size. While the CM and PD were similar between both trifunctional enzymes, NeTyrA PDT activity appeared to be over two orders of magnitude higher than the $V/K_{M\ app}$ value calculated from the parameters reported for the trifunctional TyrA from *A. fulgidus.* The main feature, however, distinguishing the PD of NeTyrA

from most other prephenate dehydrogenases characterized to date, including TyrA from *A. fulgidus*, is its cofactor specificity.

Bioinformatic studies by Jensen and colleagues on TyrA enzymes have pinpointed the residues important for cofactor specificity in the dehydrogenase-catalyzed reaction[29]. According to these studies, the key residues are found at the cofactor discriminator region located on the variable loop downstream to the GxGxxG consensus motif in the Wierrenga fingerprint (Appendix 3C) [29, 93]. According to Jensen *et al*[29], the two key residues that dictate the cofactor selectivity in TyrA proteins are the negatively charged aspartate or glutamate that dictate the specificity to NAD^+, and the positively charged Arg at the adjacent position in the sequence that is important for $NADP^+$ selectivity. Based on these studies and the amino acid sequence alignment with a number of TyrA proteins (Fig 1.8) it was predicted that the archaeal TyrA enzymes would utilize $NADP^+$ in the PD-catalyzed reaction; at the cofactor discriminator region IhTyrA and NeTyrA possess key arginine residues at positions 127 and 32, respectively, and lack the negatively charged group at the preceding position (Fig 1.8) [29]. Both TyrA PD proteins, however, displayed high activity in the presence of either $NADP^+$ or NAD^+, although with a preference for $NADP^+$; the catalytic efficiency was approximately fivefold higher with $NADP^+$ although the apparent K_M for $NADP^+$ was tenfold lower than for NAD^+ (Table 3.1 and Fig 3.3). This broad cofactor specificity is unusual for prephenate dehydrogenases that generally use NAD^+ and will be discussed further in Chapter 4. To date, the monofunctional PD from the halophilic archaea *M. mahii* is the only reported archaeal PD which also prefers $NADP^+$ as a cosubstrate (this was evaluated in crude cell extracts)[50]. Interestingly, characterization of a single-stranded DNA binding protein from *N. equitans* also revealed broad substrate specificity (in that case, the ability to bind all types of nucleic acids) suggesting a common primitive strategy for catalytic versatility in this smallest archaeon [177]. Why do the archaeal TyrA proteins

exhibit broad nucleotide cofactor specificity? It has been proposed that broad cofactor specificity is the ancestral state prior to the divergence of the NAD^+ and $NADP^+$ specific descendants which evolved in response to the mechanisms that enhanced abundance of each of the cofactors in cells [29, 49]. In accordance, the phylogeny studies place the hyperthermophillic archaeal symbionts close to the ancestral root indicating a slow rate of evolution[178]. Thus, the TyrA proteins from the archaeal symbionts appear to represent a distinct type of prephenate dehydrogenases which exhibit a primitive strategy of substrate processing.

Regulation of activity by end products

The end products of the shikimate pathway, L-phenylalanine and L-tyrosine, play an important role in the feedback regulation of CM, PD and PDT in microorganisms. Accordingly, we examined the effects of L-Phe and L-Tyr on IhTyrA and NeTyrA activities. Our studies, however, show that the CM and PD activities of TyrAs from both archaeal species were not affected by L-Tyr or by its fluorinated analog *m*-fluoro-D, L-Tyrosine which is considered a more potent inhibitor (Fig 3.4) [159]. This insensitivity to L-Tyr is unusual and contrasts most PD proteins characterized to date. This includes the *E. coli* CM-PD (Fig 3.4) and more recent examples from hyperthermophiles such as the PD activity of *A. fulgidus* and *A. aeolicus* which are inhibited by L-Tyr [52, 55, 90]. The CM and PD activities of both archaeal TyrA proteins were also not regulated by L-Phe (Fig 3.4). This is similar to other bifunctional CM-PD enzymes [97, 179], and for several monofunctional enzymes [51, 99, 180, 181], but is contrary to PD from *B. subtilis* whose activity was strongly inhibited by L-Phe[100]. The lack of regulation of CM and PD activities by L-Phe or L-Tyr has also been observed in *Z. mobilis*, as well as in the legumes *Glycine max* and *Medicago truncatula* [40, 118]. The insensitivity of PD to L-Tyr in these organisms could be explained by the existence of an additional biosynthetic pathway for L-Tyr via arogenate dehydrogenase (AD), therefore, PD may be not at the branch point leading to the production of L-Tyr or L-Phe. However, our preliminary studies with crude

arogenate preparation revealed that although IhTyrA could utilize arogenate in the dehydrogenase reaction, it did so very poorly relative to prephenate (1/200th the rate). No such ability was detected for NeTyrA. Those TyrA proteins that efficiently utilize both L-arogenate and prephenate, such as from *P. stutzeri*[51], *Z. mobilis*[182] and *P. aeruginosa*[162], do so with values of k_{cat}/K_M values for L-arogenate and prephenate that are within tenfold of each other. More thorough studies are needed in order to determine the presence of any alternative routes for L-Tyr or L-Phe biosynthesis in *I. hospitalis* and *N. equitans*.

As revealed from comparative genomic analysis, NeTyrA is predicted to retain a C-terminal ACT domain associated with allosteric regulation by end products. As expected, NeTyrA PDT was markedly inhibited by L-Phe and displayed a mixed type of non-competitive inhibition (Fig 3.5) which is indeed consistent with L-Phe binding to a site distinct from that of prephenate. This binding likely occurs at the ACT domain since neither CM nor PD activities were inhibited by L-Phe. The binding of L-Phe to the ACT domain of the enzyme caused a fivefold decrease in affinity of the enzyme for prephenate (K_i = 0.8 μM, αK_i = 4.4 μM). A similar inhibition profile has also been observed with the independently expressed PDT enzyme from *E. coli* CM-PDT [102]. The NeTyrA PDT showed no cooperative effects in its interactions with prephenate in the presence of L-Phe under our experimental conditions, which is contrary to the *E. coli* CM-PDT whose PDT activity exhibited concave upwards kinetics [183], but is similar to the engineered monofunctional PDT from *E. coli* [102]. It has been proposed for *E. coli* CM-PDT that the CM domain is responsible for the cooperative effects and for higher order L-Phe-induced oligomer formation[102]. It may be that cooperative effects were not observed for NeTyrA due to heterogeneity of the sample (see also discussion in Chapter 2).

We also noted that L-Tyr did not affect NeTyrA PDT activity at concentrations up to

1 mM. Our findings differ from those reported for the trifunctional enzyme from *A. fulgidus* whose PDT activity was activated by L-Tyr and competitively inhibited by L-Phe [(55)]. Both trifunctional enzymes are from extreme thermophiles yet they appear to adopt distinct mechanisms for efficient processing of the labile substrates. Our studies suggest that the NeTyrA is regulated solely by dehydratase activity. It may be that in the absence of L-Phe, heat-labile prephenate is preferentially channeled from the CM domain to PDT while the potent feedback inhibition of PDT by L-Phe facilitates structural changes such that prephenate is shuttled to the PD domain to yield L-Tyr. In addition, we found that IhTyrA CM and PD are not directly regulated by end products. Therefore, we speculate that in *I. hospitalis* the regulation of chorismate flux to L-Tyr occurs indirectly via PDT. Our speculation agrees with genome analysis of *I. hospitalis* that has identified gene locus Ign_0307 which is predicted to encode a distinct putative monofunctional PDT with a C-terminal ACT domain, implying allosteric regulation by L-Phe [(121)].

3.4 Summary

Characterization of TyrA enzymes from the symbiotic archaeal pair *N. equitans* and *I. hospitalis* has revealed very stable, comparably efficient enzymes with patterns of broad cofactor specificity, although with a preference for $NADP^+$, and a lack of sensitivity to L-Tyr. Additionally, NeTyrA was shown to be sensitive to regulation by L-Phe presenting a mixed type of non-competitive inhibition. The characterization confirmed that TyrA proteins from the symbiotic archaea represent uncommon members of the TyrA family. In addition to its high thermal stability, IhTyrA appeared to be remarkably stable to proteolysis and yielded a stable PD domain. The rare PD/CM/PDT organization of the trifunctional protein, the unusual cofactor specificity and L-Tyr deregulation, as well as the extreme stabilty of TyrAs from both symbionts makes these proteins attractive starting points for more detailed structural and functional investigations.

Chapter 4

Characterization of Selected Variants of *I. hospitalis* TyrA

4.0 Introduction

Expression and purification of IhTyrA, described in Chapter 2, was accompanied with a lower abundance species at a smaller molecular weight. ESI-MS and tandem MS of this ~ 30 kDa band resolved by SDS-PAGE identified a C-terminal portion of the protein for which the first 80 amino acids were missing. Additionally, limited proteolysis of IhTyrA with trypsin, described in Chapter 3, identified a stable ~ 30 kDa protein fragment; MS analysis suggested that the stable fragment belongs to the C-terminal PD domain commencing with residue 81 in the primary sequence. Sample heterogeneity from protein degradation was also reported during the expression and purification of *A. aeolicus* PD[(89)]. This impeded protein crystallography. Limited proteolysis combined with mass spectrometry, however, identified the stable C-terminal domain which was later expressed (Δ19PD) leading to the first crystal structure of a PD enzyme[(89)]. Interestingly, *H. influenzae* TyrA is a bifunctional CM-PD, yet the only reported crystal structure is of its engineered PD component that lacks the first 80 amino acid residues[(92)]. In this study we aimed to independently express the monofunctional *I. hospitalis* PD variant (Δ80CM-PD) generated by deleting the first 80 amino acids. Additionally, a conservative amino acid replacement at position 81 in IhTyrA at the junction of the mutase and dehydrogenase domains might alter the structure and/or function of either domain or the protease-sensitive region thereby reducing smaller molecular weight protein contaminants generated under our expression/purification conditions. Coincidentally, the nucleotide sequence encoding Met88 is down stream of a Shine Dalgarno sequence, which in theory allows the translation of the shorter form of the enzyme. Accordingly, the variant Met81Leu was also constructed. In this chapter we present the expression, purification and characterization of Δ80CM-PD and Met81Leu. Biophysical and biochemical characteristics of these variants are compared to the wild-type IhTyrA.

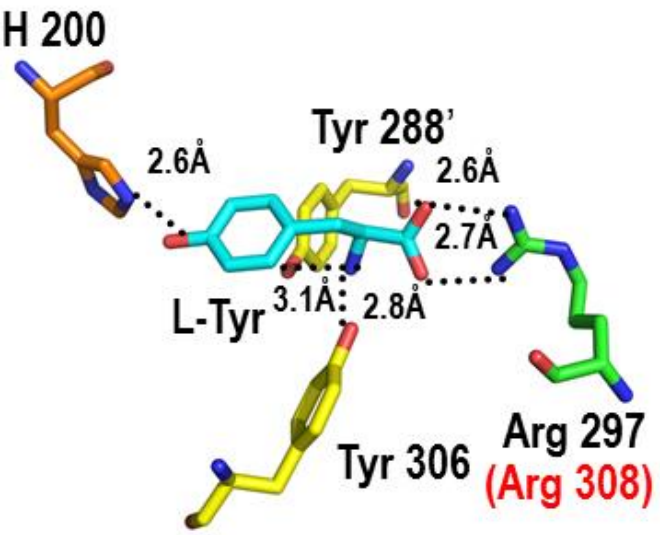
Kinetic analysis presented in Chapter 3, showed that the PD component of both NeTyrA and IhTyrA exhibited broad cofactor specificity, although with a preference for $NADP^+$, which is unregulated by L-Tyr. These characteristics are unusual as most prephenate dehydrogenases are NAD^+-specific and L-Tyr sensitive. In this Chapter we sought to identify residues that are important for cofactor specificity, for binding of prephenate and for sensitivity to L-Tyr in IhTyrA. Bioinformatic studies by Jensen and colleagues[29] demonstrated important residues that are found within the cofactor descriminator region downstream of the GxGxxG consensus motif (see Appendix 3B and Fig 1.8). According to these studies the presence of the negatively charged aspartate or glutamate is essential for NAD^+ specificity; it hydrogen bonds to the ribose near the adenine of the cofactor and repels the negatively charged phosphate of $NADP^+$. $NADP^+$-specific enzymes, however, deploy glycine, serine, threonine or alanine in place of a negatively charged aspartate or glutamate, followed by a positively charged arginine. According to Jensen *et al* [29] this reverse-in-charge allows binding of the specific cofactor. Indeed, sequence alignment (Fig 1.8) shows that NAD^+- specific dehydrogenases such as PDs from *H. influenzae*, *E. coli* and *A. aeolicus* possess aspartate or glutamate at the corresponding position (position 129 in *E. coli* CM-PD amino acid sequence) while *I. hospitalis* and the $NADP^+$-specific *Synechocystis sp.* PCC 6803 [41] have Gly126 and Ser30, respectively, followed by an arginine. In an effort to identify the residues that are critical for the cofactor specificity in *I. hospitalis* TyrA, site-specific variants were characterized: Gly126Asp, Arg127Ala/ Leu and Asn128Ala/Asp (Jensen *and colleagues*[29] indicated that aspargine is often present in species with broad cofactor specificity).

Site-directed mutagenesis studies on *E. coli* CM-PD identified Arg294 as a key residue for prephenate binding; the Arg294Gln variant had a greater than 100-fold increase in K_M for prephenate without affecting turnover of the PD-catalyzed reaction [58]. Additionally, crystallography studies of the engineered PD domain from *H. influenzae* in complex with NAD^+

and L-Tyr[92] indicated that the cationic guanidine group of Arg297 (equivalent to Arg294 in *E. coli* CM-PD) is hydrogen bonding with the side chain carboxyl group of L-Tyr (Fig 4.1) and thus may interact with the carboxyl group of prephenate. Additionally, site-directed mutagenesis studies on the PD of bifunctional CM-PD from *H. influenzae* revealed that the variant Arg297Gln, yielded an increase in K_M for prephenate of 1000-fold[95]. The arginine residue is conserved among TyrA proteins including the IhTyrA (Fig 4.1). Thus, to examine the importance of *I. hospitalis* Arg308 (equivalent to Arg294 in *E. coli* CM-PD) in prephenate binding, the variants Arg308Leu/Gln/Ala were characterized.

The crystal structure of the engineered *H. influenzae* PD identified residues Tyr288' and Tyr306 (each from adjacent monomer) that are within hydrogen bonding distance of the amine of L-Tyr (Fig 4.1) [92]. Guided by the crystal structure as well as a model of the PD domain of *E. coli* CM-PD that was based on that structure, site-directed mutagenesis studies with *E. coli* CM-PD confirmed the importance of the residues Tyr285' and Tyr303 (equivalent to Tyr288' and Tyr306 in *H. influenzae* CM-PD) in binding of L-Tyr[94]. In contrast to L-Tyr sensitive CM-PDs from *H. influenzae* and *E. coli*, *I. hospitalis* CM-PD appeared to be insensitive to L-Tyr, and the equivalent tyrosine residues that are proposed to interact with the end product are not conserved in the enzyme. In order to understand whether it is possible to introduce sensitivity to L-Tyr into IhTyrA, site-directed mutagenesis studies were employed to substitute two tyrosine residues at positions Phe299 and Ile316 (equivalent to positions Tyr288' and Tyr306 in *H. influenzae* TyrA).

A



B

```
1. I. hospitalis    HFALVSLDEAAKKLSKE-YGVDYMRYATRSFKKTLETIQKLKELS-EVIDEIQEMNEYAA 327
2. N. equitans      HLLLISYYHFARR------FPIKTASAEALYRL----SEKLLEQNPQILLDIQKEKNAK- 198
3. A. fulgidus      HFALVSMADFLRYGKE--ELKYASPIFTVLYKL----ASKIINQNWEMYFQIQK--NAE- 210
4. E. coli          HFATFAYGLHLAEENVQ-L-EQLLALSSPIYRLELAMVGKLFAQDPQLYADIIMSSERNL 314
5. H. influenzae    HFSTFANGLHLSKQPVN-L-ANLLALSSPIYRLELAMIGKLFAQDAELYADIIMDKPENL 317
6. A. aeolicus      HAVAFALVDTLIHMS-TPEV-DLFKYPGGGFKD----FTKIAKSDPIMWRDIFLENKENV 270
7. Synechocistys    VMVSAALIQACAGEKDGDILKLAQNLASSGFRD----TSKVGGGNPELGTMMATYNQRAL 237
```

Figure 4.1: (A) Crystal structure of the active site of liganded dimeric *H. influenzae* PD, and (B) the amino acid sequence alignment of the appropriate regions of the TyrA proteins. (A) PDB-2pv7[92]. The ligand L-Tyr is highlighted in light blue. Arginine 308 (in red) is the equivalent position in IhTyrA. (B) Conserved arginines (Arg308 in IhTyrA) are highlighted in red while tyrosines that are thought to be (or in case of *E. coli* CM-PD were shown to be) important are highlighted in yellow.

4.1 Experimental procedures

4.1.1 Materials

Substrates (prephenate and chorismate) were obtained as previously described by Dudzinski and Morrison[184]. L-tyrosine and *m*-fluoro-D, L-tyrosine were obtained from Sigma while NAD^+ and $NADP^+$ (Grade I) were from Roche. Concentrations of stock substrate solutions were determined using published extinction coefficients [130] and/or enzymatic end-point analysis. Oligonucleotides of standard purity were ordered from Integrated DNA Technologies. Restriction enzymes *Dpn*I, *Nde*I, *Bam*HI and *Phusion®* High-Fidelity DNA Polymerase (2.0 U/µL) were purchased from New England Biolabs Inc. The deoxy-NTP (dNTP) mixture (5 mM of each dNTP, stored at -20°C in small aliquots) was purchased from MBI Fermentas. All other chemical reagents were obtained commercially and were of the highest quality available. Purified CM-PD enzymes from *H. influenzae* and *E. coli* were prepared as described elsewhere[94, 95].

4.1.2 Source of recombinant variants of IhTyrA and wild-type enzyme

DNA mutants Met81Leu, Δ80CM-PD, Arg127Leu, Arg127Ala, Asn128Asp, Asn128Ala, Arg308Leu, Arg308Ala, Arg308Gln, Phe299Tyr and Iln316Tyr were constructed by TOP Gene Technologies. The expression plasmid pET-15b harbouring the *I. hospitalis tyrA* gene served as template DNA for the mutations. All mutants were verified by DNA sequencing.

Recombinant IhTyrA was obtained as described in Chapter 2. Recombinant *E. coli* CM-PD was expressed and purified as described previously[58] while N-terminal his-tagged *H. influenzae* CM-PD was obtained as described elsewhere[95].

4.1.3 Site-directed mutagenesis of Gly126Asp

Oligonucleotides encoding the G126D replacment (IDT Technologies):

P1 5'GTGGTGGTCACCGATAGGAACTTGGAG3' (T_m = 71 °C) and

P2 5'CTCCAAGTTCCTATCGGTGACCACCAC3' (T_m = 71 °C) were resuspended in milliQ water to a final concentration of 100 µM. Plasmid DNA of the *I.hospitalis tyrA* gene cloned into pET-15b served as template DNA for mutagenesis and was isolated from 10 mL cultures using the GeneJET™ Plasmid Miniprep Kit and their concentrations were determined spectrophotometrically at 260 nm. Site-directed mutagenesis was conducted according to the instructions supplied in the *Phusion*® site-directed mutagenesis Kit (New England Biolabs). The reaction mixtures were prepared with 15 ng of double stranded (ds) DNA template, 150 ng of each oligonucleotide primer (forward and reverse), 0.5 µL of *Phusion*® High-Fidelity DNA polymerase (2 U/µL), 10 µL of 5x *Phusion*® HF buffer, 2 µL of a 5 mM dNTP solution and 1 µL DMSO in a final volume of 50 µL. DNA polymerase was added just prior to the first denaturation cycle. PCR amplification was carried out using a GeneAmp PCR system 9700 (Applied Biosystems). The PCR conditions are listed bellow in Table 4.1

Number of Cycles	Temperature (°C)		Time
1	Denaturation	98	45 sec
35	Denaturation	98	30 sec
	Annealing	60	40 sec
	Extension	72	4 min
1	Extension	72	10 min
	Cooling	4	overnight

Table 4.1: PCR-mediated site-directed mutagenesis parameters

Upon completion of the temperature cycling, the presence of amplified DNA was confirmed by electrophoresis on a 1 % agarose gel. The methylated and hemi-methylated parental DNA was digested by incubating the PCR mixture with 10 U of *Dpn*I at 37°C overnight. *Dpn*I-treated products (5-10 µL) were transformed into *E. coli* XL10-Gold competent cells (50 µL). Briefly, purified *Dpn*I-treated PCR products were mixed with 50 µL of XL10-Gold competent cells and incubated on ice for 10 min. The cells were then incubated at 42°C for 45 s (heat-shock) and placed back on on ice for 10 min. The transformed cells were then plated on LB/agar containing 100 µg/mL ampicillin and grown overnight at 37°C. Single colony transformants were propagated in 10 mL of LB/Amp overnight at 37°C with shaking at 225 rpm and the recombinant plasmid DNA was extracted using the GeneJET^{TM} Plasmid Miniprep Kit. The concentration of plasmid DNA was determined by measuring the OD_{260}. The purified plasmid DNA samples were analyzed by sequencing at Genome Québec Innovation Center at McGill University. The resulting sequences were verified using the BLAST tool in NCBI (http://www.ncbi.nlm.nih. gov/blast) to ensure that the desired mutation was present and that no other mutations had been introduced during the PCR reaction.

4.1.4 Expression and purification of IhTyrA variants

All variants were expressed in *E. coli* BL21(DE3) cells that were harbouring plasmid pMgK and purified by Ni-NTA chromatography as described in 2.1.7. The variant Gly126Asp was derived from a 2 L cell culture, while Met81Leu and Δ80CM-PD were derived from 1 L cell cultures.

Variants Arg127Leu, Arg127Ala, Asn128Asp, Asn128Ala, Arg308Leu, Arg308Ala, Arg308Gln, Phe299Tyr and Ile316Tyr were purified after expression from 50 mL of cell culture using a “batch” Ni-NTA chromatography method. Briefly, the frozen cell pellets were resuspended in 5 mL ice cold purification buffer (50 mM Tris, 500 mM NaCl, 5% glycerol pH 8)

supplemented with 5 mM imidazole and 0.5 mM phenylmethyl sulfonyl fluoride and sonicated by 10 bursts of 10 sec each with 1 min interval on ice between each burst. Following centrifugation, each cell-free extract (about 4 mL) was then gently mixed overnight at 4°C with 500 µL of agarose Ni-NTA resin (Qiagen), and equilibrated with purification buffer (with 5 mM imidazole). The protein-resin slurry was applied onto empty bio-spin columns (Biorad). The resin was washed twice with purification buffer containing 50 mM imidazole and the bound protein was eluted with 1 mL of the same buffer but with 300 mM imidazole and stored at 4°C. The purified CM-PD variants were analyzed for protein concentration, PD activity and by SDS-PAGE as described in Chapter 2.

4.1.5 Determination of enzyme activity and inhibition by end product

PD and CM activities were determined as outlined in 2.1.13 at 80°C and pH 8 under standard reaction conditions unless otherwise stated. For the mesophilic *E. coli* and *H. influenzae* CM-PDs, assays were performed at 30°C using a 3-component buffer system of 0.1 M MES, 0.1 M 4-ethylmorpholine, 51 mM diethanolamine, at pH 7.2 supplemented with 1 mM EDTA and 1 mM DTT as described previously [26, 95] and the kinetic parameters for cofactors were determined by keeping prephenate at 0.5 mM and varying NAD^+ (0.025-1 mM) or $NADP^+$ (0.25-10 mM). Variants Phe299Tyr and Ile316Tyr were screened for inhibition by 1 mM *m*-fluoro-D,L-tyrosine at 80°C using a concentration of 0.5 mM prephenate with either of 2 mM NAD^+ or 0.5 mM $NADP^+$ as outlined in 3.1.5. Kinetic parameters for all reactions were calculated as described in 2.1.13 using substrate concentrations listed in the text.

4.1.6 Mass spectrometry

Subunit molecular weight of IhTyrA variants was determined by ESI-MS as outlined in 2.1.14. Protein samples were subjected to HPLC prior to ESI-MS analysis. Tandem MS experiments were carried out on tryptic digestions of SDS-PAGE resolved variant protein samples as outlined in 2.1.14.

4.1.7 Far-UV circular dichroism spectroscopy

Far-UV circular dichroism (CD) studies were performed on IhTyrA variants using a Jasco-815 CD spectropolarimeter as outlined in 3.1.6. Spectra were recorded from 200-260 nm at 25°C in a 0.2 cm path length cell [(52)]. For variable temperature experiments, changes in ellipticity at 222 nm (1 nm band width) were recorded from 25°C to 95°C with a ΔT of 20°C/h as described in 3.1.6.

4.1.8 Fluorescence spectroscopy

Intrinsic fluorescence of proteins (~ 5 μM monomer) was measured in 50 mM potassium phosphate buffer (pH 7.5), at ambient temperature, using the Varian Cary Eclipse Fluorescence Spectrophotometer with a 1 cm path length cell. Upon excitation at either 280 or 295 nm, emission spectra were recorded from 300–400 nm with a scan rate of 10 nm/s, excitation and emission bandwidths of 5 nm and voltage of 600 V. Emission spectra were averaged over 10 scans with a 1.0 nm data sampling interval. All spectra were baseline-corrected and the data obtained were smoothed using average smoothing function (Varian).

4.1.9 Determination of native molecular weight

4.1.9.1 Analytical size exclusion chromatography

The native molecular weights of *I. hospitalis* Δ80CM-PD and Met81Leu variants were determined at ambient temperature by using a BioRad DuoFlow FPLC system fitted with a Superdex GE-200 column (HR 10/30, Pharmacia) as outlined in 2.1.15. Protein elution was monitored at 280 nm and fractions (0.5 mL) were assayed for PD activity.

4.1.9.2 Analytical ultracentrifugation

Native molecular weight of Δ80CM-PD variant of *I. hospitalis* was investigated by sedimentation velocity experiments. Samples preparations, sedimentation data acquisition and calculation of sedimentation coefficient (s) and an average molar mass were conducted as outlined in 2.1.16.

4.1.10 Modeling of *I. hospitalis* CM and PD domains

MODELLER (V 9.11) (http://salilab.org/modeller/) was used to model the *I. hospitalis* PD monomer against the available structure of *A.aeolicus* Δ19PD (PDB-3ggg), *H. influenzae* Δ80CM-PD (PDB-2pv7), and *S. meliloti* PD (PDB-4wji). Similarly, the structure of CM from *T. thermophilus* (PDB-2d8d) was used to generate a model of the *I. hospitalis* CM domain. Protein modelling was kindly performed by Drs. D. Christendat, Francis MacManus and M. Leibovitch. Protein visualization was performed using PyMOL[(185)].

4.2 Results

4.2.1 Expression, purification and mass spectrometry analysis of IhTyrA variants Δ80CM-PD and Met81Leu

In order to determine whether the PD domain of *I. hospitalis* could yield an independently folded and functional PD enzyme as well as to provide further insight regarding the 30 kDa fragment retained in preparations of the full-length protein, the variants Δ80CM-PD and Met81Leu were expressed and purified via Ni-NTA chromatography as outlined in 4.1.4. A summary of the purification of these variants is presented in Table 4.2 and Figure 4.2.

SDS-PAGE analysis confirmed overexpression of the variants: a substantial band corresponding to the monomer size of Δ80CM-PD (~ 30 kDa) was observed in the cell lysate (Fig 4.2, lane 3). As observed for the wild-type enzyme, however, the majority of the overexpressed protein was insoluble (Fig 4.2 lanes 1-4 and Fig 2.6 lanes 1-3). A similar patern of expression was observed for the Met81Leu variant (data not shown). Nevertheless, soluble, active enzymes were detected by assaying for PD specific activity (Table 4.2) and the variants were successfully purified by Ni-NTA chromatography. Purified Δ80CM-PD appeared homogeneous showing only a ~ 30 kDa PD fragment (Fig 4.2 Lane 7) while that of of Met81Leu yielded the expected ~ 42 kDa species accompanied by a low abundance (5-10% total protein amount) of the ~ 30 kDa fragment (Fig 4.2 Lane 8). The amount of ~ 30 kDa species in the purified Met81Leu sample, however, was significantly lower than in the wild-type CM-PD sample (Fig 4.2 Lane 9). The amount of purified protein obtained from a 1 L cell culture was 3 and 4 mg for Δ80CM-PD and Met81Leu, respectively. Regrettably, concentrations of both variants above 3 mg/mL under our buffer conditions resulted in protein precipitation. The final yield of PD activity was approximately 20% (Table 4.2), comparable to that of the wild-type enzyme.

ESI-MS analysis confirmed the identity of the desired variants. Table 4.3 summarizes the expected and calculated masses for Δ80CM-PD and Met81Leu, as well as other variants expressed in this study. The analysis revealed that similar to the wild-type enzyme, the His-tagged variants lacked the first methionine at the tag's N-terminus. Interestingly, in the case of the Met81Leu variant, a degradation product at ~ 30 kDa fragment was not detected by ESI-MS (data not shown). However, tandem MS analysis of tryptic digestions of the ~ 30 kDa band from the denaturing gel of Met81Leu identified peptides from the C-terminal PD domain of the protein (data not shown).

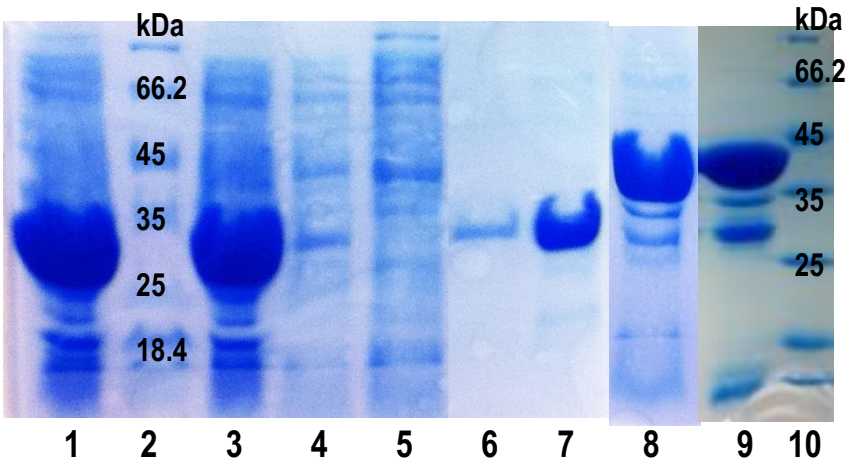


Figure 4.2: SDS-PAGE analysis of the purification of IhTyrA variant Δ80CM-PD and of the purified Met81Leu variant. Lane 1: Pellet of insoluble cellular debris solubilized in 4% SDS (from Δ80CM-PD expression), Lane 2: Protein Molecular Weight Marker, Lane 3: Cell-lysate of Δ80CM-PD expression, Lane 4: Cell-free extract of Δ80CM-PD expression Lane 5: Ni-NTA column flow through, Lane 6: Pooled Δ80CM-PD protein, Lane 7: Concentrated and stored Δ80CM-PD protein, Lane 8: Met81Leu variant purified by Ni-NTA chromatography (His-tagged), Lane 9: WT CM-PD purified by Ni-NTA (His-tagged), Lane 10: Protein Molecular Weight Marker.

Protein	Culture volume (L)	Protein amount (mg)	PD S. A. (U/mg)	Yield PD (%)
WT	2	21	6*	26
Δ80CM-PD	1	3	21	19
Met81Leu	1	4	17	20

Table 4.2: Summary of yields and specific activities of purified IhTyrA variants Δ80CM-PD and Met81Leu. Prephenate dehydrogenase activity determined at 80°C, pH 8, 0.5 mM prephenate, 2 mM NAD^+ as described in 4.1.5. *PD specific activity of purified wild-type IhTyrA fluctuated from 5-11 U/mg within similar protein purification preparations.

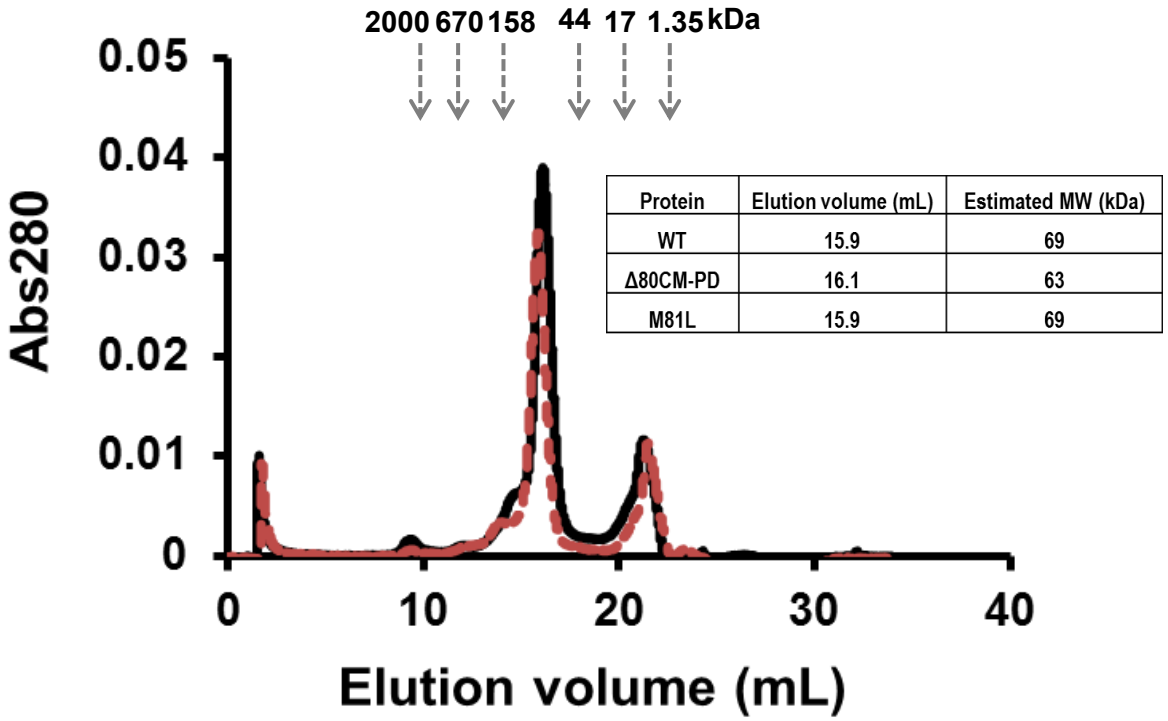
Protein	Expected Mass (Da)	Observed Mass (Da)
WT	42233	42233
Δ80CM-PD	32602	32600
Met81Leu	42215	42216
Arg127Ala	42148	42151
Arg127Leu	42190	42191
Asn128Ala	42321	42325
Asn128Asp	42334	42237
Arg308Ala	42148	42149
Arg308Gln	42205	42203
Arg308Leu	42190	ND
Phe299Tyr	42249	42250
Ile316Tyr	42283	42286
Gly126Asp	42291	42296

Table 4.3: Summary of molecular weights of purified IhTyrA and variants. Protein molecular weights were determined by LC-MS method described in 2.1.14. ND-not detected. Expected protein mass was calculated using EXPASY PeptideMass program http://web.expasy.org/peptide_mass/

4.2.2 Native molecular weight determinationof IhTyrA variants Δ80CM-PD and Met81Leu

The analysis of Met81Leu and Δ80CM-PD variants by size exclusion FPLC gave native molecular weights of 69 kDa and 63 kDa, respectively, suggesting a dimeric structure (Fig 4.3), similarly to the wild-type enzyme (see Fig 2.12). The value for the native molecular weight obtained for Δ80CM-PD variant is in good agreement with the calculated dimer mass of 65 kDa and also with the value of 57 kDa obtained from analytical ultracentrifugation sedimentation velocity experiments (Fig 4.4). Size exclusion FPLC and sedimentation velocity analysis revealed an additional lower abundance species in the Δ80CM-PD sample corresponding to molecular

weights of 111 kDa or 130 kDa respectively (Fig 4.3 and Fig 4.4). A lower abundance species corresponding to a molecular weight of 171 kDa was also observed in the Met81Leu preparation (Fig 4.3). The higher molecular weight, lower abundance species was also observed in the sample of wild-type CM-PD (see Fig 2.11 and 2.12) and we attributed this higher molecular weight species to a tetrameric form of the enzyme. Thus, analysis by size exclusion and ultracentrifugation showed that the proteins were not homogeneous.



Protein	Elution volume (mL)	Estimated MW (kDa)
WT	15.9	69
Δ80CM-PD	16.1	63
M81L	15.9	69

Figure 4.3 Size exclusion-FPLC analyses of purified IhTyrA variants Δ80CM-PD and Met81Leu. Elution profile of Δ80CM-PD is shown in black while Met81Leu shown in red. Arrows show the elution pattern of each of the indicated molecular weight standards (BioRad). SEC-FPLC was berformed as outlined in 2.1.15. Values for the wild-type enzyme were determined previously (2.2.7).

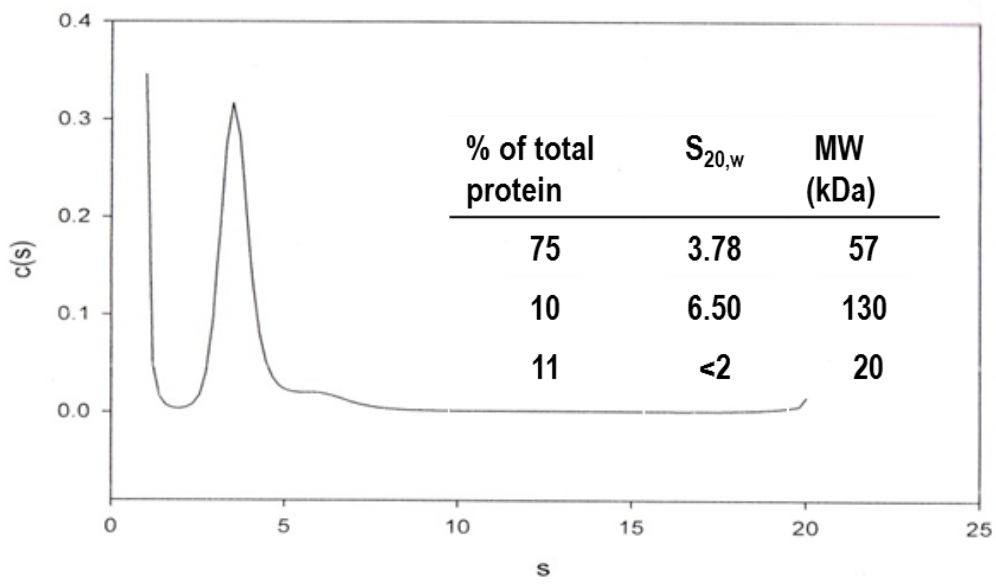


Figure 4.4: Analytical ultracentrifugation analysis of IhTyrA Δ80CM-PD variant. Protein sample at a monomer concentration of 3.1 µM, corresponded to absorbance 280 nm reading of 0.1. Values for $S_{20,W}$ and molecular weight were determined as described in 2.1.16.

4.2.3 Far-UV circular dichroism and fluorescence spectroscopy of wild-type IhTyrA and Δ80CM-PD variant

Far-UV CD spectra were recorded to determine if truncation of the mutase domain had altered the global secondary structure of the enzyme. As reported for the wild-type enzyme, the far-UV CD spectra of the Δ80CM-PD variant recorded at 25°C showed two minima at 208 nm and 222 nm indicating that both proteins possessed a significant content of α-helical structure and that truncation did not perturbed the global secondary structure (Fig 4.5).

To determine the effect of the mutase domain on the thermal stability of the bifunctional enzyme, the full-length and Δ80CM-PD proteins were analyzed by monitoring the changes in ellipticity at 222 nm when heated from 25°C to 95°C (Fig 4.6). Little change was observed with Δ80CM-PD implying that as with the full-length protein the variant was very resistant to thermal

unfolding and the T_m > 95°C under our experimental conditions. The variant retained its α- helical structure at 95°C with no formation of β-aggregates as detected by far-UV CD.

As a complementary technique to far-UV CD measurements, fluorescence emission spectroscopy was used to determine if deletion of the CM domain causes changes in tertiary structure as probed by the environment of the protein's aromatic chromophores. Emission spectra of protein samples (full-length and Δ80CM-PD variant) were recorded from 300 to 400 nm at excitation wavelengths (λ_{ex}) of 280 nm and of 290 nm. At an excitation wavelength (λ_{ex}) of 280 nm, tryptophan and tyrosine residues contribute to the proteins intrinsic fluorescence emission, while at λ_{ex} of 295 nm, the contribution is almost exclusively from tryptophan residues[(186)]. Accordingly, higher fluorescence intensities are expected at λ_{ex} of 280 than at λ_{ex} of 295. IhTyrA possesses 13 Tyr per and 5 Trp per monomer. In contrast, Δ80CM-PD possesses one less Tyr and Trp per monomer contributing to its $\lambda_{280\ nm}$ and $\lambda_{295\ nm}$ emission spectra. Fig 4.7 shows a maximal emission intensity of 337 nm at both λ_{ex} of 295 nm and 280 nm for the full-length protein and the Δ80CM-PD variant. This maximum is considerably blue shifted relative to a value of ~ 355 nm expected for full solvent exposed tryptophan residues[(187)], therefore suggesting that globally, the tryptophan residues are partially buried. As expected, higher emission intensities are observed at λ_{ex} of 280 compared to those at λ_{ex} of 295 for both the wild-type protein and the variant (Fig 4.7). Interestingly, the Δ80CM-PD exhibited a threefold decrease in fluorescence emission intensitiy at λ_{ex} of 280. However, the decrease in fluorescence emission intensity at λ_{ex} of 295 of the Δ80CM-PD variant was less pronounced. The λ_{em} max of the variant was also blue shifted but only ~ 2 nm, indicating that Trp residues are slightly more buried than in the full-length protein.

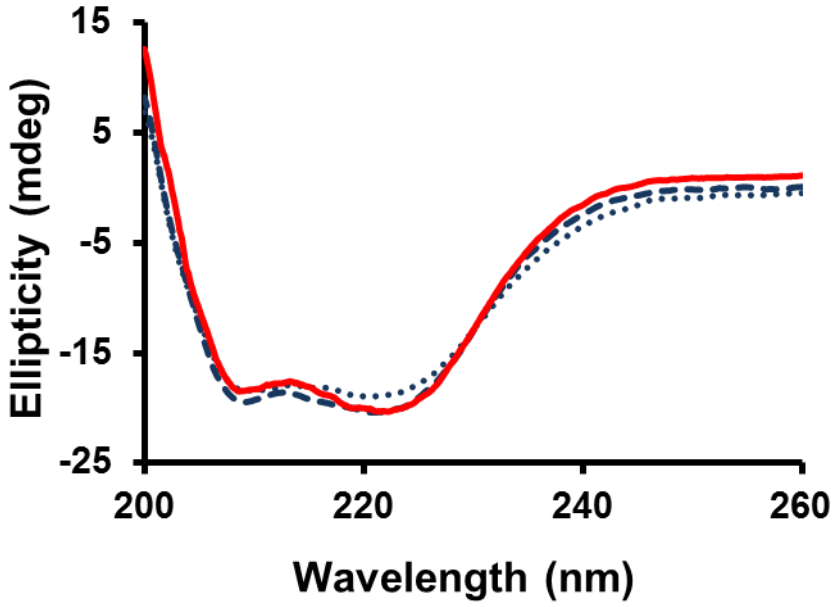


Figure 4.5: Far-UV CD spectra of wild-type IhTyrA and Δ80CM-PD variant. CD spectra of wild-type (red solid line) and Δ80CM-PD variant (blue dashed line) were recorded at 25 °C. CD spectra of the Δ80CM-PD variant were also recorded after thermal unfolding at 95°C (dotted line). The amount of protein used was approximately 0.1-0.3 mg/mL protein. The CD signal was not converted to molar ellipticity due to uncertainty in the exact concentration of protein following equilibration. Spectra are normalized to ~ 5 µM monomer.

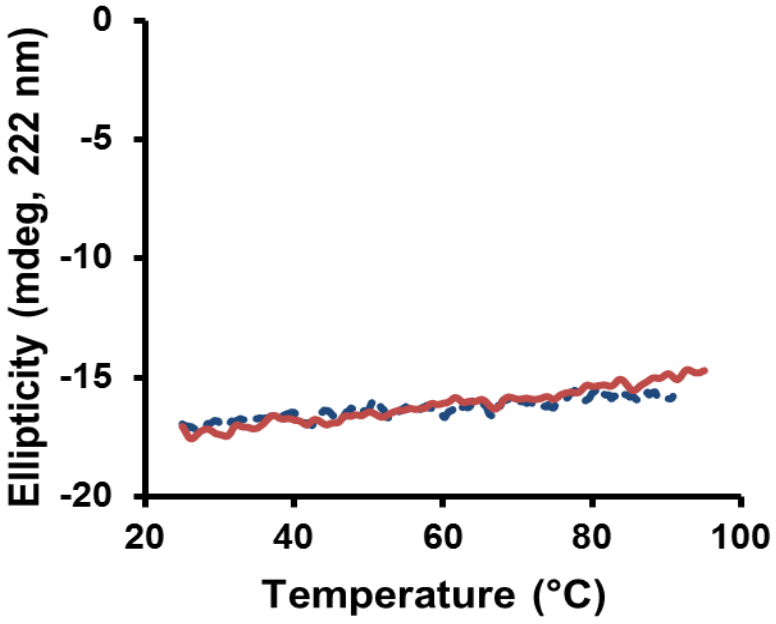


Figure 4.6: Thermal unfolding monitored by CD spectroscopy. Thermal denaturations with wild-type IhTyrA (red solid line) and Δ80CM-PD variant (dotted blue line) were carried out by following the ellipticity at 222 nm as described in 3.1.6.

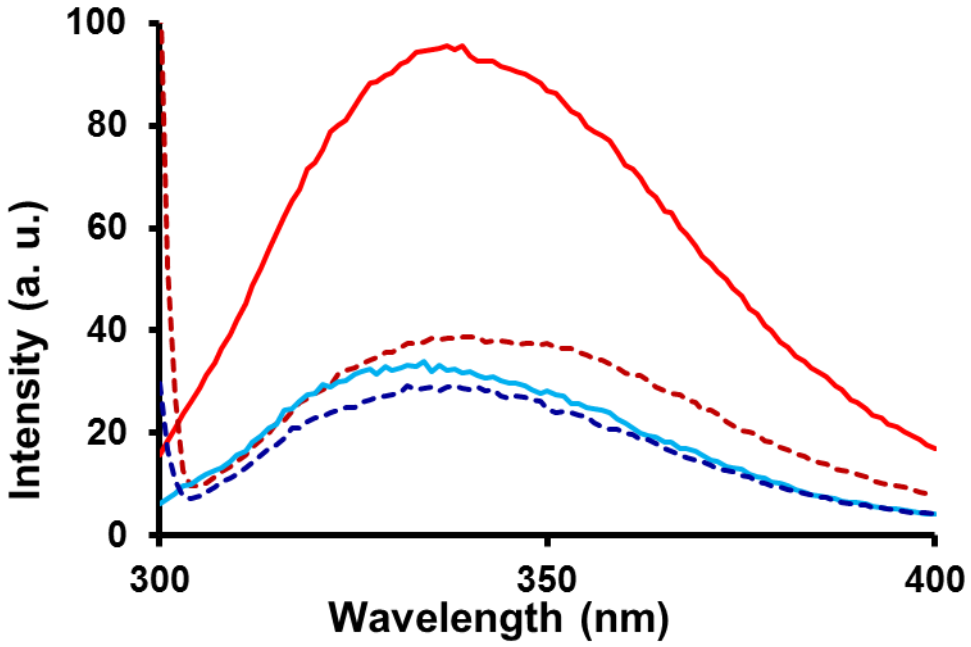


Figure 4.7: Fluorescence emission spectra of wild-type IhTyrA and Δ80CM-PD variant. Fluorescence spectroscopy analysis was performed with ~ 5 µM monomer of either wild-type CM-PD (red) or Δ80CM-PD (blue) in a buffer containing 50 mM potassium phosphate (pH 7.5) as described in 4.1.8.The excitation wavelength was set at 280 nm (solid line) or 295 nm (dotted line).

4.2.4 Kinetic characterization of IhTyrA variants Δ80CM-PD and Met81Leu

Enzyme assays with purified Δ80CM-PD and Met81Leu proteins confirmed that the variants were efficient prephenate dehydrogenases and Met81Leu appeared to be also an efficient mutase. Table 4.4 summarizes the kinetic parameters obtained for the variants and the wild-type enzyme. Kinetic characterization of Met81Leu revealed that values of k_{cat}, K_M and k_{cat}/K_M for the mutase reaction were similar to those of wild-type CM-PD. As expected, the Δ80CM-PD variant possessed no mutase activity consistent with the deletion of the enzyme's N-terminal domain. Values of k_{cat} obtained for the dehydrogenase reaction using prephenate and

NAD^+ or $NADP^+$ as substrates were similar among the variants and the wild-type enzyme. However, Δ80CM-PD exhibited a significantly higher affinity for prephenate; six- and fourfold reductions in K_M value for prephenate were observed for Δ80CM-PD in the presence of $NADP^+$ or NAD^+, respectively, when compared to the wild-type enzyme. This decrease in K_M yielded proportional increases in overall PD catalytic efficiency (k_{cat}/K_M values) of Δ80 CM-PD. There was a smaller but significant decrease in K_M values for both cofactors for Δ80CM-PD. In the case of Met81Leu, the K_M for prephenate was similar to the wild-type enzyme; however, the leucine substitution caused a pronounced reduction in the K_M value for cofactors (~ tenfold decrease in K_M for NAD^+ and >4 fold reduction in K_M for $NADP^+$). Thus, the Met81Leu replacement appeared to enhance cofactor binding without affecting the enzyme's affinity to bind other substrates or the turnover of substrates in either the CM or PD reactions.

Thermal stability studies at 95°C were performed as described in 3.1.3, and revealed that as with the wild-type enzyme, the Δ80 CM-PD variant is highly resistant to heat. As illustrated in Fig 4.8, PD specific activity of the Δ80 CM-PD variant increased about threefold during the first 1-3 h of incubation. The specific activity of wild-type PD also increased during this period although only twofold. The increase in specific activity observed for both enzymes was not due to the decrease in protein content. Although the variant appeared to be more active during the first 3 h of incubation at 95°C, it appeared less thermostable with prolonged incubation; Δ80CM-PD lost ~ 70% of its PD activity by 50 h of incubation while wild-type enzyme retained over 90% of activity within that time period. The temperature dependency of PD activity of Δ80CM-PD showed an exponential increase in reaction rate with temperature, reaching a maximum at ~ 85°C (Fig 4.9), which is below the temperature optimum of 90°C observed for the wild-type enzyme (Fig 3.2). At 90°C, the activity of Δ80CM-PD decreased by

25%. Nonetheless, the PD activity of the variant from 30° - 85°C yielded an activation energy of 47 kJ/mol which is very similar to that determined for the wild-type enzyme (Fig 4.9, inset).

	Mutase activity			Dehydrogenase activity					
Variable S	Chorismate			Prephenate			$NADP^+$ (NAD^+)		
Fixed S	------			$NADP^+$ (NAD^+)			Prephenate		
Protein	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT CM-PD	435 ± 34	33 ± 4	7.6×10^4	443 ± 105 (319 ± 35)	16 ± 3 (15 ± 2)	3.6×10^4 (4.7×10^4)	62 ± 15 (561 ± 27)	15 ± 3 (21 ± 2)	2.4×10^5 (3.7×10^4)
Met81Leu	455 ± 47	34 ± 2	7.4×10^4	463 ± 69 (238 ± 25)	14 ± 1 (13 ± 0.4)	3.0×10^4 (5.4×10^4)	14 ± 4* (45 ± 2)	14 ± 2 (12 ±0.2)	9.8×10^5 (2.7×10^5)
Δ80 CM-PD	ND	ND	ND	70 ± 9 (83 ± 13)	16 ± 1 (17 ± 1)	2.2×10^5 (2.0×10^5)	22 ± 2 (362 ± 44)	16 ± 2 (15 ± 1)	7.0×10^5 (4.2×10^4)

Table 4.4: Steady-state kinetics parameters for the reactions catalyzed by wild-type lhTyrA and variants Met81Leu and Δ80CM-PD.
Assays were performed as indicated in 2.1.13 and 4.1.5. Values were calculated from the initial rates using at least 6 substrate concentrations. CM reaction - chorismate was varied from 0.025 - 1.5 mM. PD reaction – prephenate was varied from 0.02 - 4 mM at a fixed concentration of 2 mM NAD^+ or 0.5 mM $NADP^+$. When varying the cofactor concentration: $NADP^+$ (0.015 - 4 mM) or NAD^+ (0.015 - 4 mM), prephenate was fixed at 2.5 mM. Values in parentheses are the results of assays performed with NAD^+. ND is activity not detected. For k_{cat} calculations the exact mass values used were 42, 233 (WT CM-PD), 42, 215 (Met81Leu) and 32, 602 (Δ80CM-PD).* Met81Leu - not possible to determine initial rate of PD reaction at <12 μM $NADP^+$.

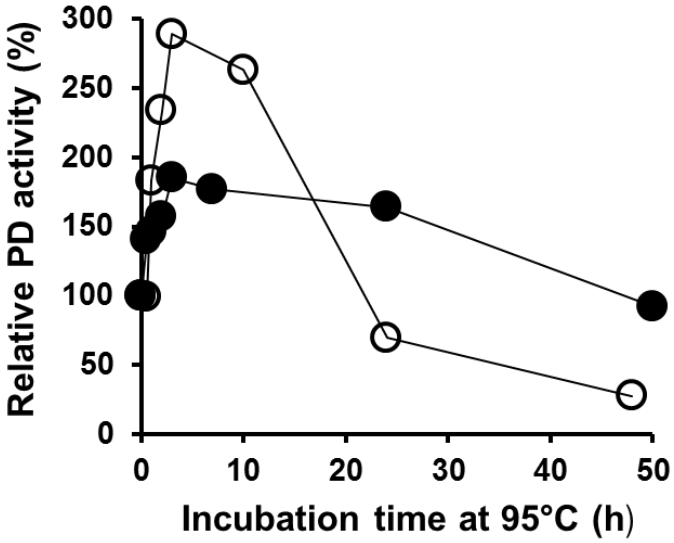


Figure 4.8: Thermal stability profile of wild-type IhTyrA and Δ80CM-PD variant. Wild-type (filled circles) and Δ80CM-PD (empty circles). Residual prephenate dehydrogenase activity was measured after incubation at 95 °C. At different time intervals during the incubation, samples were removed, centrifuged for 5 min at 4°C and residual dehydrogenase activity was determined at 80°C with 1 mM prephenate and 1 mM $NADP^+$. Protein concentration was determined after centrifugation to calculate specific activities.

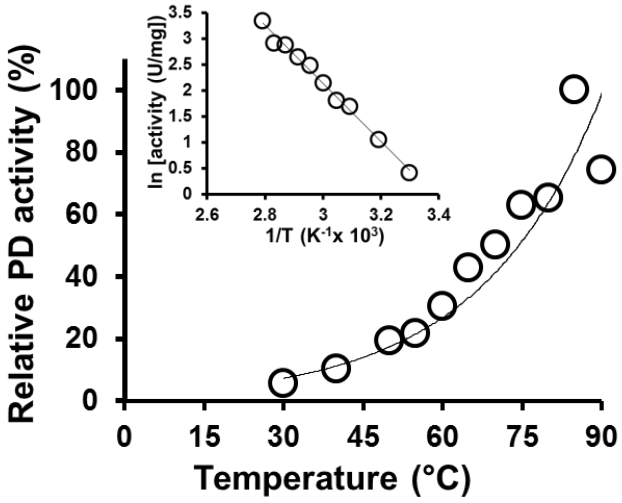


Figure 4.9: Effect of temperature on PD activity of IhTyrA Δ80CM-PD variant. PD activity was measured as described in 2.1.13. Activity at 85°C corresponds to 28 U/mg. The Arrhenius plot (inset) was generated from measurements over the range of 30° – 85°C.

4.2.5 Kinetic screening of site-specific variants to identify residues important for cofactor specificity, substrate binding and L-Tyr inhibition

Variants Arg127Ala, Arg127Leu, Asn128Ala, Asn128Asp, Arg308Leu, Arg308Gln, Arg308Ala, Phe299Tyr and Ile316Tyr along with the wild-type enzyme were expressed and purified on a small scale with Ni-NTA affinity resin while the Gly126Asp variant was expressed from a 2 L cell culture and purified by Ni-NTA chromatography as described in 4.1.4. All variants were reasonably well expressed as judged by SDS-PAGE and purified to a similar level of homogeneity as the wild-type protein (Fig 4.10). Approximately 100 - 250 μg of each variant was obtained from a 50 mL cell culture except for Arg308Leu and Arg127Leu whose amounts were lower (~30-50 μg). Purification of the Gly126Asp variant protein by Ni-NTA chromatography from a 2 L cell culture yielded 24 mg of protein. ESI-MS was used to confirm the presence of the correct amino acid substitution. As illustrated in Table 4.3, the theoretical and observed mass values of each variant protein were in good agreement. The exception was the Arg308Leu variant whose deconvoluted spectrum did not reveal any major peaks at the expected range between 20 000 and 50 000 Da (data not shown). This is likely due to the low amount of protein that was obtained through purification; the SDS-PAGE analysis, however, detected the appropriate band at ~ 42 kDa that corresponded to the expected monomer mass of the variant (Fig 4.10, A, Lane 11).

In order to investigate the importance of selected active site residues of IhTyrA for PD enzyme function, kinetic assays were performed on the wild-type and variant proteins. The analysis revealed that all variants were active in the presence of both cofactors (Tables 4.5 and 4.6). Variants Arg127Ala, Arg127Leu, Asn128Ala, Asn128Asp and Gly126Asp that were generated to assess the importance of the replaced residue in the cofactor specificity, exhibited similar specific activities with NAD^+ or $NADP^+$ implying that the replaced residues were not

critical for cofactor specificity. Moreover, kinetic analysis of Gly126Asp revealed a 13-fold increase in K_M values for both cofactors without altering cofactor preference (Table 4.6). This implied that the Gly126 is likely important in binding the cofactors rather than for specificity. To verify that the amino acid substitution of glycine to aspartate did not significantly perturb the overall protein structure we performed far-UV CD and fluorescence emission measurements on the Gly126Asp variant and the wild-type enzyme. The CD and fluorescence spectra obtained for the two proteins were nearly superimposable indicating that the amino acid replacement did not perturb the global secondary and tertiary structure of the protein (data no shown).

The importance of the conserved Arg308 in prephenate binding was evaluated by substituting the amino acid with alanine, glutamine or leucine. As indicated in Table 4.5, variants Arg308Ala and Arg308Gln exhibited a two to threefold decrease in PD activity in the presence of either of the cofactors when compared to the wild-type enzyme. The Arg308Leu exhibited a fourfold and a sixfold decrease in PD activity in presence of NAD^+ or $NADP^+$, respectively. This modest decrease in activity of the variants suggests that Arg308 is not of critical importance for prephenate binding. These findings are contrary to that observed for the CM-PD enzymes from *E. coli* and *H. influenzae* in which replacements by glutamine at the homologous positions (Arg194 and Arg197, respectively) resulted in 100-fold and 1000-fold decreases in PD activity[58, 95], respectively.

Variants Phe299Tyr and Ile316Tyr were generated to determine the possibility of introducing sensitivity to L-Tyr into IhTyrA; coordinating tyrosines are present in the corresponding positions of L-tyrosine sensitive PDs such as CM-PD from *E. coli* and *H. influenzae*. However, both variants were not inhibited by 1 mM of the fluorinated L-Tyr analog indicating that the Phe299Tyr and Ile316Tyr replacements did not introduce sensitivity to the end product (Table 4.5).

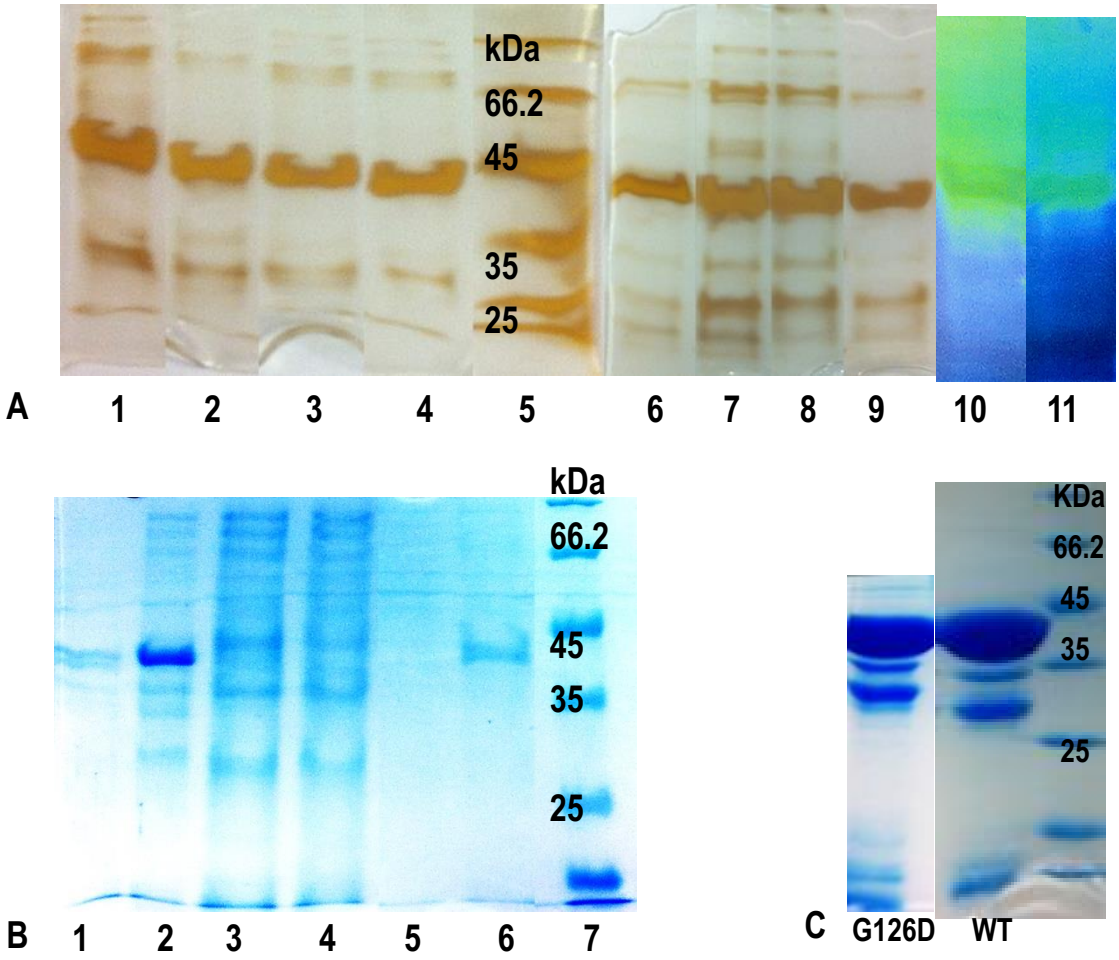


Figure 4.10: SDS-PAGE analysis of purified IhTyrA variants.

A. Small scale purified variants. Lane 1: WT, Lane 2: Arg127Ala, Lane 3: Asn128Ala, Lane 4: Arg308Ala, Lane 5: Molecular Weight Marker, Lane 6: Arg308Gln, Lane 7: Phe299Tyr, Lane 8: Ile316Tyr, Lane 9: Asn128Asp, Lane 10: Arg127Leu (modified picture), Lane 11: Arg308Leu (modified picture). Denaturing gel was silver stained to visualize proteins. Approximately 0.5-2 µg of protein was applied in each lane.

B. Purification steps of Arg308Gln. This is a representative gel for the "batch" purification method. Lane 1: Pellet of insoluble cellular debris solubilized in 4% SDS, Lane 2: Cell-lysate, Lane 3: Cell-free extract, Lane 4: Ni-NTA flow-through, , Lane 5: 30 mM imidazole wash, Lane 6: Pooled protein from Ni-NTA purification, Lane 7: Molecular Weight Marker.

C. Large scale purified Gly126Asp variant and wild-type protein. Approximately 60 µg of protein was applied in each lane.

Protein	PD activity with NAD^+		PD activity with $NADP^+$		PD activity with 1 mM DL-F-Tyr
	S.A. U/mg	% of WT S.A.	S.A. U/mg	% of WT S.A.	%
WT	9.2	100	8.8	100	100
Phe299Tyr	7.1	77	5.5	63	100
Ile316Tyr	3.5	38	3.9	44	91
Arg127Ala	11.5	125	12.5	142	
Arg127Leu	8.7	94	6.7	76	
Asn128Ala	9.9	108	9.9	113	
Asn128Asp	8.6	93	6.9	78	
Arg308Ala	4.4	48	4.4	50	
Arg308Gln	3.1	34	3.2	36	
Arg308Leu	2.2	24	1.4	16	

Table 4.5: Summary of specific activities for wild-type IhTyrA and variants. PD specific activity was determined at 80°C with NAD^+ (2 mM NAD^+, 0.5 mM prephenate) and with $NADP^+$ (0.5 mM $NADP^+$, 0.5 mM prephenate). Specific activity ± 1 mM DL-FTyr was determined with 0.5 mM $NADP^+$ and 0.5 mM prephenate.

	Mutase activity			Dehydrogenase activity					
Variable S	Chorismate			Prephenate			$NADP^+$ (NAD^+)		
Fixed S	------			$NADP^+$ (NAD^+)			Prephenate		
Protein	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
WT CM-PD	435 ± 34	33 ± 4	7.6×10^4	443 ± 105 (319 ± 35)	16 ± 3 (15 ± 2)	3.6×10^4 (4.7×10^4)	62 ± 15 (561 ± 27)	15 ± 3 (21 ± 2)	2.4×10^5 (3.7×10^4)
Gly126Asp	301 ± 34	43 ± 1.6	1.4×10^5	163 ± 38 (263 ± 13)	6 ± 0.3 (10 ± 0.2)	3.5×10^4 (3.9×10^4)	656 ± 116 (3609 ± 842)	6 ± 0.3 (7±0.5)	8.7×10^3 (1.8×10^3)

Table 4.6: Steady-state kinetics parameters for the reactions catalyzed by wild-type IhTyrA and Gly126Asp variant. Assays were performed as indicated in 2.1.13 and 4.1.5. Values were calculated from the initial rates using at least 6 substrate concentrations. CM reaction - chorismate was varied from 0.025 - 1.5 mM. PD reaction – prephenate was varied from 0.02 - 4 mM at a fixed concentration of 2 mM and 40 mM NAD^+ for WT and Gly126Asp variant, respectively or 0.5 mM $NADP^+$. When varying the cofactor concentration, $NADP^+$ (0.015 - 4 mM) or NAD^+ (0.015 - 10 mM), prephenate was fixed at 2.5 mM. Values in parentheses are the results of assays performed with NAD^+. ND is activity not detected. For k_{cat} calculations the exact mass values used were 42, 233 (WT CM-PD) and 42, 291 (Gly126Asp).

4.3 Discussion

This study reports the first heterologous expression and purification of TyrA variant proteins from *I. hospitalis.* We also provide the first functional studies on these proteins. Monofunctional PD variant Δ80CM-PD, engineered genetically by deletion of the first 80 amino acids from TyrA's mutase domain, was characterized biochemically and biophysically and compared to the wild-type bifunctional CM-PD. This comparative analysis has provided valuable information regarding the structure and function of the dehydrogenase domain separately and within the bifunctional enzyme. The characterization of single-site variant, Met81Leu has provided further insight on the spatial relationship between the mutase enzyme and the cofactor binding site within the dehydrogenase domain. Our studies also extended to single-site variants Gly126Asp, Arg127Ala, Arg127Leu, Asn128Ala and Asn128Asp, Arg308Leu, Arg308Gln, Arg308Ala, Phe299Tyr and Ile316Tyr which caused amino acid replacements at positions predicted to play a role in either cofactor specificity, substrate binding or sensitivity to L-Tyr.

Deletion of the mutase domain yields a dimeric, active monofunctional dehydrogenase

The Δ80CM-PD variant was expressed and purified to apparent homogeneity by Ni-NTA affinity chromatography to yield an intact domain devoid of small molecular weight fragments as determined by SDS-PAGE (Fig 4.2, lane 7) and ES-MS analysis (Table 4.3). As expected, the enzyme lacked mutase activity. However, it retained full dehydrogenase activity showing a preference for the cofactor $NADP^+$, similar to the wild-type enzyme. The independently expressed PD domain also behaved as the full-length protein exhibiting maximum activity at high temperatures (Fig 4.9).

Size exclusion chromatography and analytical ultracentrifugation analysis confirmed that Δ80CM-PD was dimeric in solution indicating that the mutase domain was not essential for the higher order structure of the bifunctional enzyme (Fig 4.3 and Fig 4.4). This is in accord with previous studies on the mesophilic *H. influenzae* CM-PD and its monofunctional Δ80CM-PD variant which also revealed an active dimeric enzyme[92, 95]. The average molecular weight of Δ80 CM-PD determined from sedimentation velocity experiments was less than the value predicted from the primary sequence (Fig 4.4 and 3.10), indicating that the variant was somewhat more compactly folded than the full-length protein. This finding is also in accord with the results of the limited proteolysis experiments and the expression profile of a 30 kDa fragment during protein expression/purification (Fig 2.10, 3.9, and 3.10) which supports the idea of a more tightly folded independent PD domain. This is, however, in contrast with SEC-FPLC analysis, which suggests that the variant is less compact than the full-length protein (Fig 4.3).

The examples of engineered monofunctional PDs provided by *I. hospitalis* Δ80CM-PD (in this study) and of Δ80CM-PD of *H. influenza* [92, 95] are more successful than that reported by Ganem and colleagues. Their efforts to generate competent CM and PD fragments of *E. coli* CM-PD by deletion of the first 93 or 96 residues of the bifunctional enzyme yielded constructs that were less active and much less stable than the full-length protein[188] .

Far-UV CD spectroscopy revealed that as with the full-length protein, the Δ80CM-PD variant exhibited considerable α-helical content and its global secondary structure was retained even at high temperatures. Our studies agree with those on *H. influenzae* CM-PD and its Δ80CM-PD variant which also exhibited similar helical structure at ambient temperature. The melting temperatures (T_m) for both full-length and Δ80CM-PD proteins exceeded 95°C, a behavior which is consistent with hyperthermophilic proteins. Similar profiles using far-UV CD spectroscopy were obtained for hyperthermophilic PD from *A. aeolicus* and its Δ19PD variant [52].

Structural differences between full-length CM-PD and Δ80CM-PD

Changes in the shape and tertiary structure of the enzyme caused by the truncation might be reflected in the protein's intrinsic fluorescence emission, which is sensitive to the environment of aromatic chromophores. Our data show that IhTyrA and the Δ80CM-PD variant demonstrated pronounced differences in fluorescence emission intensities implying that there were tertiary structural differences. A model of the *I. hospitalis* PD domain that was generated based on the crystal structure of *H. influenzae* Δ80CM-PD (~ 25% sequence identity), shows the side chains of three tryptophan residues (Trp143, Trp235 and Trp344) per monomer are more exposed while that of Trp155 is partially buried (Fig 4.12). (Trp64 is located in the mutase domain and therefore is not found in the model). The emission maximum at λ_{ex} of 295 nm (specific for Trp), was 337 nm reflecting that on average, the tryptophan residues are partially buried [(187)]. Thus the model is only in moderate agreement with the data. The overall fluorescence emission spectra of full-length CM-PD, after excitation at 280 nm (reflecting the contribution of Trp and Tyr), displayed a ~ threefold increase in quantum yield relative to Δ80CM-PD (Fig 4.7). This is not surprising since the full-length protein has one additional tryptophan and an additional tyrosine residue in the mutase domain. While the fluorescence emission for the full-length protein at λ_{ex} of 280 nm was considerably higher than at λ_{ex} of 295 nm, for Δ80CM-PD, surprisingly the quantum yields upon excitation at these two wavelengths were almost superimposable. The 3D model of the *I. hospitalis* PD domain revealed that Trp155 is located in close proximity (~ 5 Å) to Tyr98, and Trp235 is ~ 5 Å from Tyr234. Thus, it may be that these tyrosine residues are more effectively quenched by the nearby tryptophan residues in the Δ80CM-PD protein. Additionally, the PD model showed that nine of the twelve tyrosine residues are surface exposed (Fig 4.11) and their fluorescence emission is likely quenched by

interactions with the HPO_4^- ion in the phosphate buffer[186]. Therefore, the marked differences in fluorescence intensities between the full-length and the Δ80 CM-PD proteins upon excitation at 280 nm could also be explained by a change in the solvent exposure of some of the tyrosines with the deletion of the mutase domain.

Worth noting in the model of the PD, Trp143 is within 5 Å of the cofactor docked in the active site (Fig 4.12). We are currently evaluating the accuracy of the model and if fluorescence quenching can be exploited to monitor cofactor binding in the wild-type enzyme and selected variants (for example Met81Leu, discussed below).

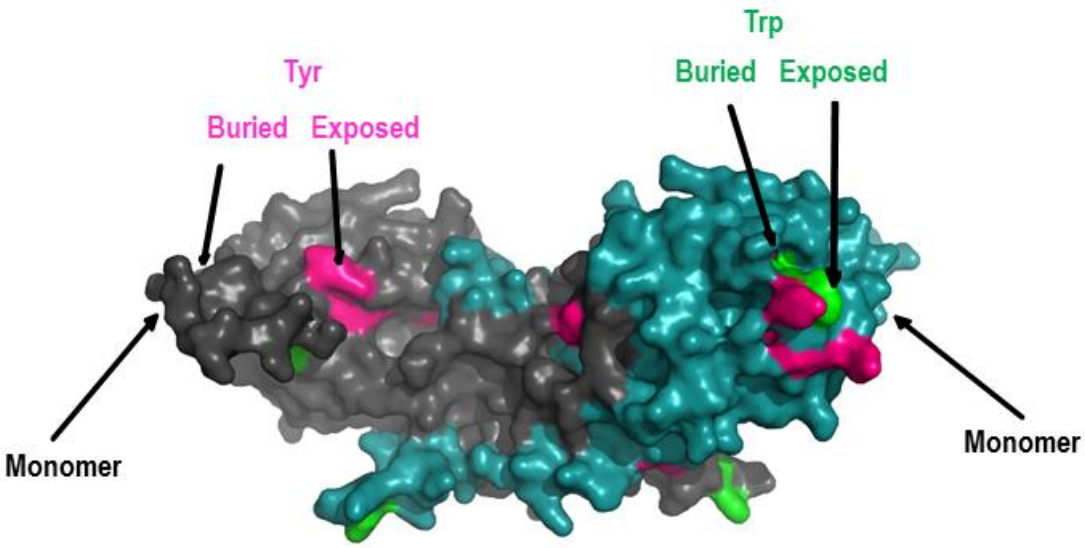


Figure 4.11: Surface accessibility of tryptophan and tyrosine residues in the model of dimeric PD domain of IhTyrA. Model of dimeric PD Domain of IhTyrA based on *H. influenzae* 3D structure: PDB-2pv7 [92, 94]. Surface-exposed tryptophan residues are shown in green and surface-exposed tyrosine residues are shown in pink. There are 4 tryptophan and 12 tyrosine residues per monomer. The two monomers of the dimer are shown in blue and gray. Picture was generated using PyMol.

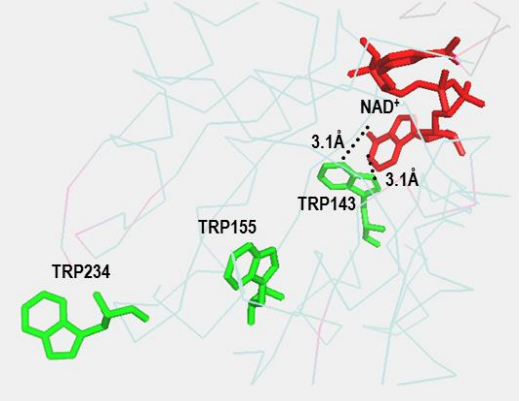


Figure 4.12: Proposed interaction of NAD^+ with Trp143 in model of PD domain of IhTyrA. Model of *I. hospitalis* PD domain based on *H. influenzae* 3D Structure: PDB-2pv7 [92, 94]. NAD^+ is highlighted in red, tryptophan residues are highlighted in green. Three out of four trytophans are shown. Trp344 is located far from the active site and therefore is not shown in the picture.

Truncation contributes to a catalytically more effective but less thermally stable domain

Changes in tertiary structure of the protein, monitored by fluorescence emission at room temperature, are also likely reflected in functional changes in the protein. Kinetic analysis revealed that *I. hospitalis* Δ80 CM-PD was more catalytically efficient than the full-length protein, exhibiting six and threefold decreases in K_M for prephenate and $NADP^+$, respectively (Table 4.5). Additionally, thermal stability studies showed that PD activities of both full-length and Δ80CM-PD were activated at high temperatures. During the first 3 h of incubation at 95°C the monofunctional variant achieved maximum activity that was twice that of the bifunctional enzyme. However, Δ80CM-PD lost most of its activity when incubated at this temperature for longer than 24 h, in contrast to the full-length protein which retained almost full activity (Fig 4.8). Additionally, it was observed that the maximal PD activity achieved by *I. hospitalis* Δ80CM-PD at 85°C was 5°C lower than the optimum obtained for full-length PD, indicating a reduced resistance to heat (Fig 4.9). Thus, deletion of the N-terminal domain of the protein appears to compromise the tertiary structure of the protein at high temperatures.

Although there are no structures available for bifunctional CM-PDs, it is likely that there is a close structural association between the mutase and dehydrogenase domains in such a way that would afford protection against thermally-induced unfolding and inactivation. Thus, our data suggest that deletion of the CM domain favors PD activity while impacting protein stability at higher temperatures.

Arnold *et al* through directed evolution studies on Cytochrome P450 enzymes, indicated that mutations in a protein's active site to enhance function are often destabilizing since they alter functional features that are also essential for maintaining stable structures such as buried charges and cavities[(189)]. Accordingly, we hypothesize that deletion of the mutase domain in *I. hospitalis* CM-PD may have exposed internal charges and/or a cavity, likely at the interface of the CM and PD domains. Examing fluorescence emission in the presence of external quenching agents may provide some additional information as has been performed for *A. aeolicus* Δ19PD[(52)] .

Amino acid replacement in the mutase domain affects cofactor binding

The Met81Leu variant was characterized to further probe the origin of the 30 kDa fragment observed during the protein expression and purification. The amount of 30 kDa fragment co-purifying wih the Met81Leu variant was proportionally lower than observed in the wild-type enzyme but not eliminated (Fig 4.2). Thus, the shorter form of the enzyme is unlikely the result of an additional protein translation site from that supplied by our *E. coli* expression vector, although this might be the case in the host organism. Met81Leu resides in the C-terminal region of the CM domain upstream of the consensus sequence for nucleotide binding. It is likely that the leucine replacement perturbs the structure of the protein in the region near the protease recognition site. Kinetic analysis of the variant shows that none of the kinetic

parameters of the CM and PD-catalyzed reactions are affected by the replacement except the K_M for the cofactor, which is significantly reduced: tenfold for NAD^+ and over fourfold for $NADP^+$ (the later measurement limited by our spectrophotometric assay). A model of the CM domain of *I. hospitalis* (Appendix 7A) displays this methionine at the protein's C-terminal helix directed away from the mutase active site. It is therefore possible that this non-polar residue could be interacting with, and alter the hydrophobic packing of a region close to the $NADP^+$ binding site within the PD domain. Suprisingly, there have been very few variant proteins characterized, which carry amino acid replacements only in the mutase portion of a bifunctional CM-PD TyrA[(57)]. This is the first mutase variant characterized that shows an effect solely on cofactor binding.

In summary, our studies are the first to provide insights on the spatial and functional relationship between the two catalytic activities in *I. hospitalis* CM-PD and reveal that PD activity can be contained within a stable monofunctional domain. This independent expression contributes to the PD activity however compromises protein thermal stability.

A survey of the importance of active site residues in ligand binding and selectivity in IhTyrA

Our work on IhTyrA was expanded to identify amino acid residues that may be important for cofactor specificity, binding of prephenate and/or (in)sensitivity to L-Tyr. Towards this goal a total of ten single-site variant proteins were characterized, and thus represent the first site-directed mutagenesis studies on *I. hospitalis* CM-PD. Amino acid substitutions were designed according to 3D homology models based on the available crystal structures of bacterial PD, as well as amino acid sequence alignments with other TyrA proteins. All variants, purified by Ni-NTA chromatography exhibited yields and levels of purity comparable to the wild-type protein.

Driven by our kinetic studies in which IhTyrA exhibited a preference for $NADP^+$ in the dehydrogenase reaction, we were intrigued by the possibility of changing the protein's preference towards NAD^+ by identifying and replacing the critical residues which may confer specificity. IhTyrA appeared to be very stable and unregulated by the end product L-Tyr, thus representing a potential candidate for applications in biotechnology through overproduction of L-Tyr; changing cofactor specificity would be advantageous since NAD^+ is more stable and much cheaper than $NADP^+$ [190]. Bioinformatics studies on TyrA proteins by Jensen and colleagues [29] has pinpointed residues likely important for cofactor discrimination that are found downstream of the GxGxxG consensus motif (see Appendix 3B and Fig 1.8). According to these studies, as well as functional studies on other dehydrogenases[191, 192], specificity to NAD^+ is thought to be conferred by a single, negatively charged residue, aspartate or glutamate since it repels the negatively charged phosphate group of $NADP^+$. The crystal structure of the PD domain from *H. influenzae* revealed that the side chain of aspartate at position 131 is within hydrogen bonding distance of the diol of ribose near the adenine of NAD^+ (Fig 4.13). The crystal structure of *A. aeolicus* Δ19PD also revealed a homologous aspartate at position 62 interacting with the cofactor[89]. Accordingly, our functional studies had shown that both bacterial enzymes are NAD^+-specific PDs (see Appendix 8A and [52, 95]. Our efforts to introduce the negative charge at the equivalent position of *I. hospitalis* PD by substituting the corresponding Gly126 to aspartate, however, did not switch cofactor preference. Instead, the variant Gly126Asp exhibited a 13-fold decrease in apparent affinity for both cofactors suggesting the importance of Gly126 in facilitating the binding of either of the two cofactors. The model of *I. hospitalis* PD based on the crystal structure of *H. influenzae* PD showed that the backbone amine group of Gly126 is within hydrogen bonding distance of the hydroxyl side chain of Tyr104 (Appendix 5A). Thus, it may be that replacing Gly126 with the much larger and negatively charged aspartate perturbed the

cofactor binding site through steric repulsion. Our results contrast with studies on alcohol dehydrogenase from *Lactobacillus brevis* which demonstrated that a single substitution Gly37Asp shifted cofactor dependence from $NADP^+$ to NAD^+ [192] by a factor of 300. Generally, attempts to alter specificity from $NADP^+$ to NAD^+ are often not successful and more success had been achieved with the reverse change, from NAD^+ to $NADP^+$-specificity[190]. In those cases, site-specific substitutions were made such that the negatively charged aspartate or glutamate residue was removed and additionally, a positively charged arginine was introduced at the adjacent position, thus reversing the charge in the cofactor binding site[191, 193]. Crystallography studies on $NADP^+$-dependent AD from *Synechocystis* sp. PCC 6803 showed that the side chain of Arg31 was hydrogen bonding with a phosphate oxygen of $NADP^+$ (Appendix 6A)[41]. Common to $NADP^+$-specific enzymes, the preceeding position is occupied by Ser30 instead of the negatively charged aspartate (or glutamate) which is typical of NAD^+-dependent dehydrogenases such as the equivalent Asp131 in *H. influenzae* PD (see sequence alignment of PDs in Fig 1.8). In order to evaluate whether PD reaction could be abolished in the presence of $NADP^+$, Arg127 (equivalent to Arg31 in *Synechocystis* sp. PCC 6803 AD) was substituted to alanine and leucine. However, both Arg127Leu and Arg127Ala variants exhibited PD activity comparable to the wild-type enzyme in the presence of either cofactor implying that the arginine does not play an important role in cofactor preference or binding in the archaeal enzyme. Furthermore, replacing Asn128 (the following position in the primary sequence) to either alanine or to the anionic aspartate had little effect on the activity in the presence of either cofactor, implying that aspargine does not play a role in cofactor preference (asparagine was previously noted by Jensen *et al*[29] in the cofactor discriminator region of TyrA proteins predicted to exhibit broad cofactor specificity). Our studies also disagree with the model of *I. hospitalis* PD based on the crystal structures of PD from *S. meliloti.* In this structure it has been demonstrated that Arg127 and Asn128 are within

hydrogen bonding distance of the phosphate oxygen of $NADP^+$ (Fig 4.14). Amino acid sequence identity between the two enzymes is relatively low (~ 17%), however, which may have contributed to the inaccuracy of the model. Our studies reveal that residues Gly126, Arg127 and Asn128 are not critical for cofactor specificity as thought. We suggest that the cofactor specificity is likely conferred by a combination of residues rather than by a single amino acid in IhTyrA.

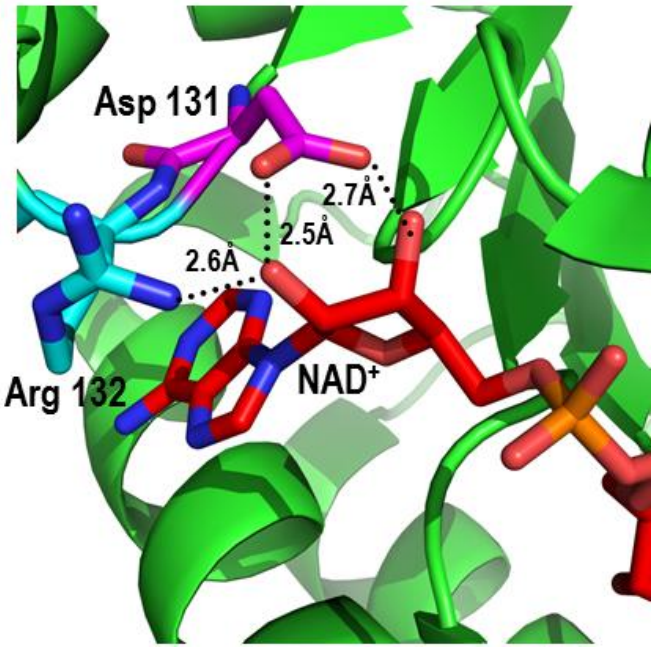


Figure 4.13: Interactions of NAD^+ with Asp131 in the crystal structure of dimeric *H. influenzae* PD. PDB-2pv7 [92]. NAD^+ is highlighted in red. Asp131 and Arg132 are highlighted in purple and blue, respectively.

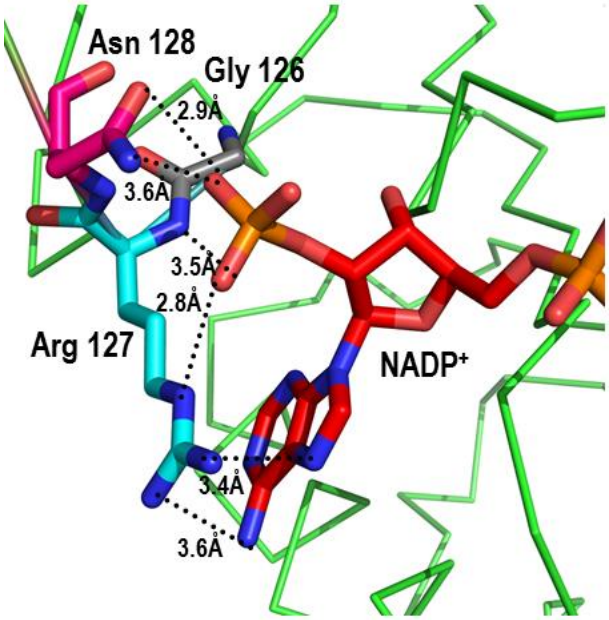


Figure 4.14: Model of the NADP$^+$-binding site of *I. hospitalis* PD. Model of *I. hopitalis* PD complexed with NADP$^+$ (red) based on the *S. meliloti* PD crystal structure (PDB-4wji). Gly126, Arg127 and Asn128 are highlighted in gray, blue and pink, respectively.

A similar conclusion can be drawn from our studies on single replacement variants of a conserved arginine residue, assigned to Arg308 in IhTyrA that plays a role in binding prephenate binding. In contrast with site-directed mutagenesis studies on *E. coli* and *H. influenzae* PDs [(58, 95)], in which the arginine at equivalent positions was deemed essential for prephenate binding, our studies revealed that Arg308 is not critical for prephenate binding in *I. hospitalis* PD. The replacement of cationic arginine with alanine or glutamine reduced PD activity in IhTyrA by only two to threefold compared to reductions in k_{cat}/K_M activity by 100 to 1000-fold by the glutamine substitutions in *E. coli* and *H. influenzae* CM-PD [(95, 96)]. Currently, there are no crystal structures

of PD bound with prephenate, although there are published structures by Christendat, Turnbull and colleagues [(90)] of *A. aeolicus* Δ19PD in complex with HPP and NADH and L-Tyr and NAD^+, and of the structure by Wilson *et al* [(92)] of *H. influenzae* Δ80CM-PDs with bound L-Tyr and NAD^+. Both structures illustrate that the conserved arginine (Arg297 in *H. influenzae* PD or Arg250 in *A. aeolicus* Δ19PD) is hydrogen bonding via its guanidinium group with the side chain carboxyl group of L-Tyr (see Fig 1.9 and 1.10). Although, the model of *I. hospitalis* PD based on the crystal structure of *H. influenzae* PD clearly shows the corresponding Arg308 within hydrogen bond distance of the L-Tyr side chain carboxylate (Appendix 9A), the backbone residues of Arg308 are clashing with Phe299. Interestingly, site-directed mutagenesis studies on Δ19PD from *A. aeolicus* suggest that Lys246' from an adjacent subunit rather than Arg250 plays a more significant role in binding prephenate and that the effect is additive with Arg250 (Arg250Ala/Lys246Ala yields a 200-fold increase in K_M for prephenate[(96)]). Moreover Legrand and colleagues[(41)] have hypothesized that in the AD from *Synechocystis* sp. PCC 6803, cationic residues near the active site play no direct role in prephenate binding but rather help to guide the substrate towards the active site. Whether this is the case for *I. hospitalis* PD awaits further mutagenesis studies.

Our efforts to introduce sensitivity to L-Tyr in *I. hospitalis* CM-PD revealed that single-site replacements to tyrosine at positions Phe299 and Ile316 were not essential for the enzyme's interactions with L-Tyr. This conclusion was drawn from kinetic data (Table 4.5) that showed that variants Phe299Tyr and Ile316Tyr exhibited PD activities that were similar in the presence or absence of the fluorinated L-Tyr analog. This analog was shown to be a more potent inhibitor in tyrosine-sensitive PD proteins [(159)]. Coordinating tyrosine residues are found at equivalent positions in the PD domains of tyrosine-sensitive bifunctional CM-PD proteins including those from *E. coli* and *H. influenzae,* as well as in the tyrosine-sensitive archaeal

trifunctional TyrA from *A. fulgidus* (See sequence alignment in Fig 1.8)[92]. Site-directed mutagenesis studies by Turnbull and colleagues on *E. coli* CM-PD [98] guided by a modeled structure of the active site revealed that these tyrosine residues (Tyr285' and Tyr303), each from adjacent monomers, are essential for binding L-Tyr through its amine group. However, the coordinating tyrosines are not conserved in monofunctional TyrA proteins such as the L-Tyr-sensitive PDs from *A. aeolicus* and *S. mutans*, as well as the *Synechocystis sp.* PCC 6803 enzyme which is unregulated by L-Tyr (See alignment in Fig 1.8). In these PD proteins as well as in the unregulated CM-PD from *I. hospitalis,* in the equivalent position to 285 in *E. coli,* tyrosine is replaced with phenylalanine while tyrosine in position 303 is replaced by a different residue in each PD. Moreover, structural studies on Δ19PD from *A. aeolicus* have demonstrated an alternate orientation of the bound L-Tyr with its amine group interacting directly with the backbone carbonyl of Thr152[90]. It is not too surprizing that the introduction of tyrosines in IhTyrA, at positions equivalent to those coordinating L-Tyr in *H. influenzae* PD, did not confer sensitivity to the end product. Our studies, presented in Chapters 2 and 3 suggest that *I. hospitalis* CM-PD along with *N. equitans* PD-CM-PDT may encompass a distinct subfamily of dehydrogenases with regard to their deregulation by L-Tyr and co-substrate specificity. Complementary crystallography studies would be beneficial to validate the differences within the PDs from the symbiotic archaeal pair and with other characterized PD proteins in terms of their active site geometries. Such studies would help to identify residues that play a role in substrate and cofactor selectivity and in conferring resistance to L-Tyr.

Chapter 5

Summary and Future Directions

5.0 Summary and future directions

This work reports the first production and characterization of trifunctional PD-CM-PDT from the smallest characterized hyperthermophilc archaeon *N. equitans* and bifunctional CM-PD from its archaeal host, crenarchaeon *I. hospitalis*. Hexa-histidine tagged proteins were expressed in *E. coli* and purified by affinity chromatography. Characterization of the recombinant archaeal TyrA proteins demonstrated that both the proteins exhibited remarkable thermal stablility and the bifunctional enzyme from *I. hospitalis* appeared to be also highly stable to proteolysis. Kinetic studies showed that the archaeal enzymes were insensitive to regulation by the end product L-Tyr and possessed unusual broad cofactor specificity with preference to $NADP^+$ in the PD reaction. Thus our studies provide evidence that PDs from the unique symbiotic archaeal pair encompass a distinct subfamily of dehydrogenases with regard to their regulation and cofactor specificity. We also showed through kinetic studies that the PDT component of the trifunctional enzyme was regulated by the end product L-Phe in a manner consistent with the combination at a site distinct from that of prephenate.

Thermal stability studies showed that PD activities of both archaeal TyrA proteins increased about two-fold during the first 1-3 hours of incubation at 95°C. Therefore, it would be important to determine the kinetic parameters of the heat treated protein sample. Generating temperature conditions as similar as possible to the native environment of these hyperthermophilic enzymes will keep them in the optimal conformation for establishing their maximal efficiency.

Additionally, these studies revealed that heating the *N. equitans* TyrA sample at 95°C for 1 h assisted in removing many contaminating *E. coli* proteins but not the chaperones SlyD and DnaK. Our studies also showed that chaperone SlyD can be separated from the NeTyrA

by size exclusion FPLC. Thus, if somebody would aim to purify further the trifunctional enzyme one could perform a heat-step on the cell-free extract at 95°C for one hour to remove most of the *E. coli* proteins, followed by Ni-NTA and by SEC-FPLC to remove the chaperone slyD. This would yield a NeTyrA sample which will still contain the chaperone DnaK (which is not separated from the archaeal protein by the methods mentioned above). It was shown elsewhere that DnaK contamination could be reduced by washing the fusion protein bound to the purification resin with a solution that contains 5 mM MgATP and soluble denatured *E. coli* proteins[(194)]. Thus, it may be that using this solution in the washes during Ni-NTA chromatography would be beneficial in removal of DnaK from the NeTyrA sample. However, taking into account the poor expression of soluble NeTyrA in *E. coli* BL21(DE3) cells, as well as the purification steps proposed above in which significant amounts of the protein of interest are removed along with contaminants, one should process a large scale expression (> 20 L of cell culture). It is likely that the intrinsic properties of the gene/protein product such as high pI, large size, AT richness of the genome will not allow heterologous overexpression of the protein. A pBAD vector could be utilized in which protein production is under very tight control.

Our studies on the bifunctional TyrA from *I. hospitalis* provide the first insights on the spatial and functional relationship between CM and PD activities. Through mass spectrometry and mutagenesis studies (representing the first mutagenesis studies on the archaeal enzyme) we demonstrated that the PD domain can be independently isolated and expressed yielding a stable and active monofunctional enzyme. We found that isolation of the PD domain contributes to PD activity by increasing the apparent binding affinity for prephenate but lowers the protein stability.

One of our goals was to perform crystallography studies on the homogeneous preparation of the archaeal protein. Of all our preparations the only nearly homogeneous

sample was of the independently expressed PD domain of *I. hospitalis* CM-PD (Δ80CM-PD). Thus, Δ80CM-PD is an interesting candidate for crystallography studies as an example of a distinct TyrA protein due to its broad cofactor specificity and deregulation by L-Tyr. As a future work, we propose to execute crystallography studies on Δ80CM-PD in the presence of its substrates: prephenate and $NAD(P)^+$. Unfortunatelly, our studies revealed that the archaeal TyrA proteins precipitate at concentrations above 4 mg/mL. Considering that a common protein concentration for crystallography studies is ~10 mg/mL[(195)], crystallization of Δ80CM-PD would be challenging. Different buffer conditions should be screened to increase the soluble protein concentration. Additionally, a Lys79Gln variant should be produced to increase the homogeneity of the full-length protein sample. An additional protease cleavage site should also be examined (through MS and site-directed mutagenesis) which appears to yield a fragment of about 35 kDa on SDS-PAGE. Modelling studies may help in determing the accessibility of that site. Additionally, modelling has suggested that intrinsic fluorescence emission will allow the development of a thermodynamic binding assay for substrates and that must be explored.

Site-directed mutagenesis studies on selected residues in *I. hospitalis* CM-PD suggest that cofactor specificity, prephenate binding and sensitivity to L-Tyr, each are conferred by a combination of residues rather than by single amino acids in the PD domain. As future directions to identify those residues it may be worthwhile to try a directed evolution approach and/or a semi-rational approach.

Our studies have suggested that the archaeal TyrAs are excellent candidates for applications in biotechnology through overproduction of L-Tyr because of their high stability and deregulation by L-Tyr. Moreover, TyrA proteins serve as an excellent model for comparative enzymology studies. Thus, our studies on the unusual archaeal TyrAs, which include the rare fusion of *N. equitans* PD-CM-PDT domains, provide a starting point to yield a wealth of

information on the structure and function of multifunctional proteins and the evolution of enzyme activity within this metabolic pathway.

References

1. Kaneko, M., Hwang, E. I., Ohnishi, Y., and Horinouchi, S. (2003) Heterologous production of flavanones in Escherichia coli: potential for combinatorial biosynthesis of flavonoids in bacteria, *Journal of Industrial Microbiolgy and Biotechnolgy 30*, 456-461.
2. Meganathan, R. (2001) Ubiquinone biosynthesis in microorganism, *FEMS Microbiology Letters 203*, 131-139.
3. Moller, B. L., and Conn, E. E. (1979) The biosynthesis of cyanogenic glucosides in higher plants. N-hydroxytyrosine as an intermediate in the biosynthesis of dhurrin by Sorghum bicolor (Linn) Moench, *The Journal of Biological Chemistry 254*, 8575-8583.
4. Memelink, J. (2004) Putting the opium in poppy to sleep, *Nature Biotechnology 22*, 1526-1527.
5. Furst, P., and Stehle, P. (2004) What are the essential elements needed for the determination of amino acid requirements in humans?, *The Journal of Nutrition 134*, 1558S-1565S.
6. Coggins, J. R., Abell, C., Evans, L. B., Frederickson, M., Robinson, D. A., Roszak, A. W., and Lapthorn, A. P. (2003) Experiences with the shikimate-pathway enzymes as targets for rational drug design, *Biochemical Society Transactions 31*, 548-552.
7. Haslam, E. (1974) *The Shikimate Pathway*, Wiley, New York.
8. Pérez, E., Rubio, M. B., Cardoza, R. E., Gutiérrez, S., Bettiol, W., Monte, E., and Hermosa, R. (2015) The importance of chorismate mutase in the biocontrol potential of Trichoderma parareesei, *Frontiers in Microbiology 6*, 1181.
9. Cabrera-Valladares, N., Martínez, A., Piñero, S., Lagunas-Muñoz, V. H., Tinoco, R., de Anda, R., Vázquez-Duhalt, R., Bolívar, F., and Gosset, G. (2006) Expression of the melA gene from Rhizobium etli CFN42 in Escherichia coli and characterization of the encoded tyrosinase, *Enzyme and Microbial Technology 38*, 772-779.
10. Bourke, S. L., and Kohn, J. (2003) Polymers derived from the amino acid L-tyrosine: polycarbonates, polyarylates and copolymers with poly(ethylene glycol), *Advanced Drug Delivery Reviews 55*, 447-466.
11. Ikram, U.-H., and Ali, S. (2002) Microbiological transformation of L -tyrosine to 3,4-dihydroxyphenyl L -alanine (L -Dopa) by a mutant strain of Aspergillus oryzae UV-7, *Current Microbiolgy 45*, 88-93.
12. Rodriguez, A., Kildegaard, K. R., Li, M., Borodina, I., and Nielsen, J. (2015) Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis, *Metabolic Engineering 31*, 181-188.
13. Santos, C. N. S., Xiao, W., and Stephanopoulos, G. (2012) Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in Escherichia coli, *Proceedings of the National Academy of Sciences of the United States of America 109*, 13538-13543.
14. Chavez-Bejar, M. I., Baez-Viveros, J. L., Martinez, A., Bolivar, F., and Gosset, G. (2012) Biotechnological production of L-tyrosine and derived compounds, *Process Biochemistry 47*, 1017-1026.
15. Davis, B. D. (1955) Intermediates in amino acid biosynthesis, *Advances in Enzymology and Related Subjects of Biochemistry 16*, 247-312.

16. Gibson, F., and Pittard, J. (1968) Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microorganisms, *Bacteriological Reviews 32*, 465-492.
17. Amrhein, N., Deus, B., Gehrke, P., and Steinrucken, H. C. (1980) The site of the inhibition of the shikimate pathway by glyphosate: II. Interference of glyphosate with chorismate formation in vivo and in vitro, *Plant Physiology 66*, 830-834.
18. Schönbrunn, E., Eschenburg, S., Shuttleworth, W. A., Schloss, J. V., Amrhein, N., Evans, J. N. S., and Kabsch, W. (2001) Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail, *Proceedings of the National Academy of Sciences of the United States of America 98*, 1376-1380.
19. Dayan, J., and Sprinson, D. B. (1971) Enzyme alterations in tyrosine and phenylalanine auxotrophs of Salmonella typhimurium, *The Biochemical Journal 108*, 1174-1180.
20. Dosselaere, F., and Vanderleyden, J. (2001) A metabolic node in action: chorismate-utilizing enzymes in microorganisms, *Critical Reviews in Microbiology 27*, 75-131.
21. Berry, A., Dodge, T. C., Pepsin, M., and Weyler, W. (2002) Application of metabolic engineering to improve both the production and use of biotech indigo, *Journal of Industrial Microbiolgy and Biotechnolgy 28*, 127-133.
22. Schmid, J. W., Mauch, K., Reuss, M., Gilles, E. D., and Kremling, A. (2004) Metabolic design based on a coupled gene expression—metabolic network model of tryptophan production in Escherichia coli, *Metabolic Engineering 6*, 364-377.
23. Han, G. H., Bang, S. E., Babu, B. K., Chang, M., Shin, H.-J., and Kim, S. W. (2011) Bio-indigo production in two different fermentation systems using recombinant Escherichia coli cells harboring a flavin-containing monooxygenase gene (fmo), *Process Biochemistry 46*, 788-791.
24. Bonvin, J. (2008) Biochemical and biophysical characterization of a prephenate dehydrogenase from the hyperthermophilic bacterium Aquifex Aeolicus, Concordia University, Montreal, CA.
25. Haslam, E. (1993) *Shikimic Acid : Metabolism and Metabolites*, Wiley, Chichester; New York.
26. Turnbull, J., Cleland, W. W., and Morrison, J. F. (1990) Chorismate mutase-prephenate dehydrogenase from Escherichia coli. 1. Kinetic characterization of the dehydrogenase reaction by use of alternative substrates, *Biochemistry 29*, 10245-10254.
27. Powell, J. T., and Morrison, J. F. (1978) The purification and properties of the aspartate aminotransferase and aromatic-amino-acid aminotransferase from Escherichia coli, *European Journal of Biochemistry / FEBS 87*, 391-400.
28. Stenmark, S. L., Pierson, D. L., Glover, G. I., and Jensen, R. A. (1974) Blue-green bacteria synthesize L-tyrosine by the pretyrosine pathway, *Nature 247*, 290-292.
29. Bonner, C. A., Disz, T., Hwang, K., Song, J., Vonstein, V., Overbeek, R., and Jensen, R. A. (2008) Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities, *Microbiology and Molecular Biology Reviews : MMBR 72*, 13-53.
30. Bonner, C., and Jensen, R. (1987) Prephenate aminotransferase, *Methods in Enzymology 142*, 479-487.
31. Rippert, P., and Matringe, M. (2002) Molecular and biochemical characterization of an Arabidopsis thaliana arogenate dehydrogenase with two highly similar and active protein domains, *Plant Molecular Biology 48*, 361-368.
32. Yamada, T., Matsuda, F., Kasai, K., Fukuoka, S., Kitamura, K., Tozawa, Y., Miyagawa, H., and Wakasa, K. (2008) Mutation of a rice gene encoding a phenylalanine biosynthetic

enzyme results in accumulation of phenylalanine and tryptophan, *The Plant Cell 20*, 1316-1329.
33. Jung, E., Zamir, L. O., and Jensen, R. A. (1986) Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate, *Proceedings of the National Academy of Sciences of the United States of America 83*, 7231-7235.
34. Siehl, D. L., and Conn, E. E. (1998) Kinetic and regulatory properties of arogenate dehydratase in seedlings of Sorghum bicolor (L.) Moench, *Archives of Biochemistry and Biophysics 260*, 822-829.
35. Patel, N., Pierson, D. L., and Jensen, R. A. (1977) Dual enzymatic routes to L-tyrosine and L-phenylalanine via pretyrosine in Pseudomonas aeruginosa, *The Journal of Biological Chemistry 252*, 5839-5846.
36. Fazel, A. M., and Jensen, R. A. (1979) Obligatory biosynthesis of L-tyrosine via the pretyrosine branchlet in coryneform bacteria, *Journal of Bacteriology 138*, 805-815.
37. Hall, G. C., Flick, M. B., Gherna, R. L., and Jensen, R. A. (1982) Biochemical diversity for biosynthesis of aromatic amino acids among the cyanobacteria, *Journal of Bacteriology 149*, 65-78.
38. Keller, B., Keller, E., and Lingens, F. (1985) Arogenate dehydrogenase from *Streptomyces phaeochromogenes*. Purification and properties, *Biological Chemistry Hoppe Seyler 366*, 1063-1066.
39. Mayer, E., Waldner-Sander, S., Keller, B., Keller, E., and Lingens, F. (1985) Purification of arogenate dehydrogenase from Phenylobacterium immobile, *FEBS Letters 179*, 208-212.
40. Zhao, G., Xia, T., Ingram, L. O., and Jensen, R. A. (1993) An allosterically insensitive class of cyclohexadienyl dehydrogenase from Zymomonas mobilis, *European journal of Biochemistry / FEBS 212*, 157-165.
41. Legrand, P., Dumas, R., Seux, M., Rippert, P., Ravelli, R., Ferrer, J. L., and Matringe, M. (2006) Biochemical characterization and crystal structure of Synechocystis arogenate dehydrogenase provide insights into catalytic reaction, *Structure 14*, 767-776.
42. Herrmann, K. M. (1995) The shikimate pathway as an entry to aromatic secondary metabolism, *Plant Physiology 107*, 7-12.
43. Gosset, G. (2009) Production of aromatic compounds in bacteria, *Current opinion in Biotechnology 20*, 651-658.
44. Grinter, N. J. (1998) Developing an L-phenylalanine process, *ChemTech 28*, 33-37.
45. Yagasaki, M., and Hashimoto, S.-i. (2008) Synthesis and application of dipeptides; current status and perspectives, *Applied Microbiology and Biotechnology 81*, 13-22.
46. Leuchtenberger, W., Huthmacher, K., and Drauz, K. (2005) Biotechnological production of amino acids and derivatives: current status and prospects, *Applied Microbiology and Biotechnology 69*, 1-8.
47. Patnaik, R., Zolandz, R. R., Green, D. A., and Kraynie, D. F. (2008) L-tyrosine production by recombinant escherichia coli: fermentation optimization and recovery, *Biotechnology Bioengineering 99*, 741-752.
48. Rippert, P., Scimemi, C., Dubald, M., and Matringe, M. (2004) Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance, *Plant Physiology 134*, 92-100.
49. Bonner, C. A., Jensen, R. A., Gander, J. E., and Keyhani, N. O. (2004) A core catalytic domain of the TyrA protein family: arogenate dehydrogenase from Synechocystis, *The Biochemical Journal 382*, 279-291.

50. Fischer, R. S., Bonner, C. A., Boone, D. R., and Jensen R. A. (1993) Clues from a halophilic methanogen about aromatic amino acid biosynthesis in archaebacteria, *Archives in Microbiology 160,* 440-446.
51. Xie, G., Bonner, C. A., and Jensen, R. A. (2000) Cyclohexadienyl dehydrogenase from Pseudomonas stutzeri exemplifies a widespread type of tyrosine-pathway dehydrogenase in the TyrA protein family, *Comparative Biochemistry and Physiology. Toxicology & Pharmacology : CBP 125*, 65-83.
52. Bonvin, J., Aponte, R. A., Marcantonio, M., Singh, S., Christendat, D., and Turnbull, J. L. (2006) Biochemical characterization of prephenate dehydrogenase from the hyperthermophilic bacterium Aquifex aeolicus, *Protein Science : a publication of the Protein Society 15*, 1417-1432.
53. Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont, *Nature 417*, 63-67.
54. Cock, J. M., Sterck, L., Rouze, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J. M., Badger, J. H., Beszteri, B., Billiau, K., Bonnet, E., Bothwell, J. H., Bowler, C., Boyen, C., Brownlee, C., Carrano, C. J., Charrier, B., Cho, G. Y., Coelho, S. M., Collen, J., Corre, E., Da Silva, C., Delage, L., Delaroque, N., Dittami, S. M., Doulbeau, S., Elias, M., Farnham, G., Gachon, C. M., Gschloessl, B., Heesch, S., Jabbari, K., Jubin, C., Kawai, H., Kimura, K., Kloareg, B., Kupper, F. C., Lang, D., Le Bail, A., Leblanc, C., Lerouge, P., Lohr, M., Lopez, P. J., Martens, C., Maumus, F., Michel, G., Miranda-Saavedra, D., Morales, J., Moreau, H., Motomura, T., Nagasato, C., Napoli, C. A., Nelson, D. R., Nyvall-Collen, P., Peters, A. F., Pommier, C., Potin, P., Poulain, J., Quesneville, H., Read, B., Rensing, S. A., Ritter, A., Rousvoal, S., Samanta, M., Samson, G., Schroeder, D. C., Segurens, B., Strittmatter, M., Tonon, T., Tregear, J. W., Valentin, K., von Dassow, P., Yamagishi, T., Van de Peer, Y., and Wincker, P. (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae, *Nature 465*, 617-621.
55. Lim, S., Springstead, J. R., Yu, M., Bartkowski, W., Schroder, I., and Monbouquette, H. G. (2009) Characterization of a key trifunctional enzyme for aromatic amino acid biosynthesis in Archaeoglobus fulgidus, *Extremophiles : life under extreme conditions 13*, 191-198.
56. Liberles, J. S., Thorolfsson, M., and Martinez, A. (2005) Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism, *Amino Acids 28*, 1-12.
57. Christendat, D., Saridakis, V. C., and Turnbull, J. L. (1998) Use of site-directed mutagenesis to identify residues specific for each reaction catalyzed by chorismate mutase-prephenate dehydrogenase from Escherichia coli, *Biochemistry 37*, 15703-15712.
58. Christendat, D., and Turnbull, J. L. (1999) Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from Escherichia coli, *Biochemistry 38*, 4782-4793.
59. Turnbull, J., Cleland, W. W., and Morrison, J. F. (1991) pH dependency of the reactions catalyzed by chorismate mutase-prephenate dehydrogenase from Escherichia coli, *Biochemistry 30*, 7777-7782.
60. Turnbull, J., and Morrison, J. F. (1990) Chorismate mutase-prephenate dehydrogenase from Escherichia coli. 2. Evidence for two different active sites, *Biochemistry 29*, 10255-10261.

61. Strater, N., Schnappauf, G., Braus, G., and Lipscomb, W. N. (1997) Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures, *Structure 5*, 1437-1452.
62. Andrews, P. R., Smith, G. D., and Young, I. G. (1973) Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate, *Biochemistry 12*, 3492-3498.
63. Gorisch, H., and Lingens, F. (1974) Chorismate mutase from Streptomyces. Purification, properties, and subunit structure of the enzyme from Streptomyces aureofaciens Tu 24, *Biochemistry 13*, 3790-3794.
64. Copley, S. D., and Knowles, J. R. (1987) The conformational equilibrium of chorismate in solution: implications for the mechanism of he non-enzymic and the enzyme-catalyzed rearrangement of chorismate to prephenate, *Journal of the American Chemical Society 109*, 5008-5013.
65. Guilford, W. J., Copley, S. D., and Knowles, J. R. (1987) *Journal of the American Chemical Society 109*, 5013-5019.
66. Andrews, P. R., and Heyde, E. (1979) A common active site model for catalysis by chorismate mutase--prephenate dehydrogenase, *Journal of Theoretical Biology 78*, 393-403.
67. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Bertchold, G. A., and Knowles, J. R. (1984) Stereocheistry of the rearrangement of chorismate to prephenate: chorismate involves a chair transition state, *Journal of the American Chemical Society 106*, 2701-2703.
68. Chao, H. S., and Berchtold, G. A. (1982) Inhibition of chorismate mutase activity of chorismate mutase-prephenate dehydrogenase from Aerobacter aerogenes, *Biochemistry 21*, 2778-2781.
69. Delany, J. J., Padykula, R. E., and Bertchold, G. A. (1992) Uncatalyzed and chorismate mutase catalyzed Claisen rearrangements of 5,6-dihydrochorismate and 6-oxa-5,6-dihydrochorismate, *Journal of the American Chemical Society 114*, 1394-1397.
70. Bartlett, P. A., Nakagawa, Y., Johnson, C. R., Reich, S. H., and Luis, A. (1988) Chorismate mutase inhibitors:synthesis and evaluation of some potential transition state analogs, *Journal of Organic Chemistry 53*, 3195-3210.
71. Turnbull, J., Morrison, J. F., and Cleland, W. W. (1991) Kinetic studies on chorismate mutase-prephenate dehydrogenase from Escherichia coli: models for the feedback inhibition of prephenate dehydrogenase by L-tyrosine, *Biochemistry 30*, 7783-7788.
72. Xue, Y., Lipscomb, W. N., Graf, R., Schnappauf, G., and Braus, G. (1994) The crystal structure of allosteric chorismate mutase at 2.2-A resolution, *Proceedings of the National Academy of Sciences of the United States of America 91*, 10814-10818.
73. Helmstaedt, K., Heinrich, G., Merkl, R., and Braus, G. H. (2004) Chorismate mutase of Thermus thermophilus is a monofunctional AroH class enzyme inhibited by tyrosine, *Archives of Microbiology 181*, 195-203.
74. Krengel, U., Dey, R., Sasso, S., Okvist, M., Ramakrishnan, C., and Kast, P. (2006) Preliminary X-ray crystallographic analysis of the secreted chorismate mutase from Mycobacterium tuberculosis: a tricky crystallization problem solved, *Acta crystallographica. Section F, Structural Biology and Crystallization Communications 62*, 441-445.
75. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of Escherichia coli chorismate mutase, *Journal of the American Chemical Society 117*, 3627-3628.

76. Calhoun, D. H., Bonner, C. A., Gu, W., Xie, G., and Jensen, R. A. (2001) The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from Salmonella typhimurium and Pseudomonas aeruginosa, *Genome Biology 2*, Research 0030.
77. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) Crystal structures of the monofunctional chorismate mutase from Bacillus subtilis and its complex with a transition state analog, *Proceedings of the National Academy of Sciences of the United States of America 90*, 8600-8603.
78. Okvist, M., Dey, R., Sasso, S., Grahan, E., Kast, P., and Krengel, U. (2006) 1.6 A crystal structure of the secreted chorismate mutase from Mycobacterium tuberculosis: novel fold topology revealed, *Journal of Molecular Biology 357*, 1483-1499.
79. Xu, H., Yang, C., Chen, L., Kataeva, I. A., Tempel, W., Lee, D., Habel, J. E., Nguyen, D., Pflugrath, J. W., Ferrara, J. D., Arendall, W. B. r., Richardson, J. S., Richardson, D. C., Liu, Z. J., Newton, M. G., Rose, J. B., and Wang, B. C. (2005) Away from the edge II: in-house Se-SAS phasing with chromium radiation., *Acta Crystallographica D Biological Crystallography 61*, 960-966.
80. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) Site-directed mutagenesis of monofunctional chorismate mutase engineered from the E. coli P-protein, *Bioorganic & Medicinal Chemistry 4*, 1015-1020.
81. Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) Analysis of Active Site Residues in Escherichia coli Chorismate Mutase by Site-Directed Mutagenesis, *Journal of the American Chemical Society 118*, 1789-1790.
82. Ishida, T. (2010) Effects of point mutation on enzymatic activity: correlation between protein electronic structure and motion in chorismate mutase reaction, *Journal of the American Chemical Society 132*, 7104-7118.
83. Lever, G., Cole, D. J., Lonsdale, R., Ranaghan, K. E., Wales, D. J., Mulholland, A. J., Skylaris, C. K., and Payne, M. C. (2014) Large-scale density functional theory transition state searching in enzymes, *The Journal of Physical Chemistry Letters 5*, 3614-3619.
84. Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems, *Angewandte Chemie 48*, 1198-1229.
85. Hermes, J. D., Tipton, P. A., Fisher, M. A., O'Leary, M. H., Morrison, J. F., and Cleland, W. W. (1984) Mechanisms of enzymatic and acid-catalyzed decarboxylations of prephenate, *Biochemistry 23*, 6263-6275.
86. Sampathkumar, P., and Morrison, J. (1982) Chorismate mutase-prephenate dehydrogenase from *Escherichia coli.* Kinetic mechanism of the prephenate dehydrogenase reaction, *Biochimica Biophysica Acta 702*, 212-219.
87. Heyde, E., and Morrison, J. F. (1978) Kinetic studies on the reactions catalyzed by chorismate mutase-prephenate dehydrogenase from Aerobacter aerogenes, *Biochemistry 17*, 1573-1580.
88. Christendat, D., and Turnbull, J. (1996) Identification of active site residues of chorismate mutase-prephenate dehydrogenase from Escherichia coli, *Biochemistry 35*, 4468-4479.
89. Sun, W., Singh, S., Zhang, R., Turnbull, J. L., and Christendat, D. (2006) Crystal structure of prephenate dehydrogenase from Aquifex aeolicus. Insights into the catalytic mechanism, *The Journal of Biological Chemistry 281*, 12919-12928.
90. Sun, W., Shahinas, D., Bonvin, J., Hou, W., Kimber, M. S., Turnbull, J., and Christendat, D. (2009) The crystal structure of Aquifex aeolicus prephenate dehydrogenase reveals the mode of tyrosine inhibition, *The Journal of Biological Chemistry 284*, 13223-13232.

91. Ku, H. K., Do, N. H., Song, J. S., Choi, S., Yeon, S. H., Shin, M. H., Kim, K. J., Park, S. R., Park, I. Y., Kim, S. K., and Lee, S. J. (2011) Crystal structure of prephenate dehydrogenase from Streptococcus mutans, *International Journal of Biological Macromolecules 49*, 761-766.
92. Chiu, H. J., Abdubek, P., Astakhova, T., Axelrod, H. L., Carlton, D., Clayton, T., Das, D., Deller, M. C., Duan, L., Feuerhelm, J., Grant, J. C., Grzechnik, A., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kozbial, P., Krishna, S. S., Kumar, A., Marciano, D., McMullan, D., Miller, M. D., Morse, A. T., Nigoghossian, E., Okach, L., Reyes, R., Tien, H. J., Trame, C. B., van den Bedem, H., Weekes, D., Xu, Q., Hodgson, K. O., Wooley, J., Elsliger, M. A., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2010) The structure of Haemophilus influenzae prephenate dehydrogenase suggests unique features of bifunctional TyrA enzymes, *Acta Crystallographica. Section F, Structural biology and crystallization communications 66*, 1317-1325.
93. Wierenga, R. K., Terpstra, P., and Hol, W. G. (1986) Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint, *Journal of Molecular Biology 187*, 101-107.
94. Hassounah, S. (2009) Characterizing residues involved in feedback regulation of a tyrosine biosynthetic enzyme, Concordia University, Montreal, CA.
95. Quashie, P. (2010) Characterization of mono- and bifunctional forms of bacterial TyrA, Concordia University, Montreal, CA.
96. Hotz, N. (2010) Active site studies of a thermophilic dehydrogenase, Concordia University, Montreal, CA.
97. Christopherson, R. I. (1985) Chorismate mutase-prephenate dehydrogenase from Escherichia coli: cooperative effects and inhibition by L-tyrosine, *Archives of Biochemistry and Biophysics 240*, 646-654.
98. Hudson, G. S., Howlett, G. J., and Davidson, B. E. (1983) The binding of tyrosine and NAD+ to chorismate mutase/prephenate dehydrogenase from Escherichia coli K12 and the effects of these ligands on the activity and self-association of the enzyme. Analysis in terms of a model, *The Journal of Biological Chemistry 258*, 3114-3120.
99. Song, J., Bonner, C. A., Wolinsky, M., and Jensen, R. A. (2005) The TyrA family of aromatic-pathway dehydrogenases in phylogenetic context, *BMC Biology 3*, 1-30.
100. Champney, W. S., and Jensen, R. A. (1970) The enzymology of prephenate dehydrogenase in Bacillus subtilis, *The Journal of Biological Chemistry 245*, 3763-3770.
101. Knaggs, A. R. (2001) The biosynthesis of shikimate metabolites, *Natural Product Reports 18*, 334-355.
102. Zhang, S., Pohnert, G., Kongsaeree, P., Wilson, D. B., Clardy, J., and Ganem, B. (1998) Chorismate mutase-prephenate dehydratase from Escherichia coli. Study of catalytic and regulatory domains using genetically engineered proteins, *The Journal of Biological Chemistry 273*, 6248-6253.
103. Zhang, S., Wilson, D. B., and Ganem, B. (2000) Probing the catalytic mechanism of prephenate dehydratase by site-directed mutagenesis of the Escherichia coli P-protein dehydratase domain, *Biochemistry 39*, 4722-4728.
104. Gething, M. J., Davidson, B. E., and Dopheide, T. A. (1976) Chorismate mutase/prephenate dehydratase from Escherichia coli K12. 1. The effect of NaCl and its use in a new purification involving affinity chromatography on sepharosyl-phenylalanine, *European Journal of Biochemistry / FEBS 71*, 317-325.
105. Fischer, R., and Jensen, R. (1987) Prephenate dehydratase (monofunctional), *Methods in Enzymology 142*, 507-512.

106. Davidson, B. E. (1987) Chorismate mutase-prephenate dehydratase from Escherichia coli, *Methods in Enzymology 142*, 432-439.
107. Kleeb, A. C., Kast, P., and Hilvert, D. (2006) A monofunctional and thermostable prephenate dehydratase from the archaeon Methanocaldococcus jannaschii, *Biochemistry 45*, 14101-14110.
108. Pierson, D. L., and Jensen, R. A. (1974) Metabolic interlock: Control of an interconvertible prephenate dehydratase by hydrophobic amino acids in Bacillus subtilis, *Journal of Molecular Biology 90*, 563-579.
109. Prakash, P., Pathak, N., and Hasnain, S. E. (2005) pheA (Rv3838c) of Mycobacterium tuberculosis encodes an allosterically regulated monofunctional prephenate dehydratase that requires both catalytic and regulatory domains for optimum activity, *The Journal of Biological Chemistry 280*, 20666-20671.
110. Gu, W., Williams, D. S., Aldrich, H. C., Xie, G., Gabriel, D. W., and Jensen, R. A. (1997) The aroQ and pheA domains of the bifunctional P-protein from Xanthomonas campestris in a context of genomic comparison, *Microbial & Comparative Genomics 2*, 141-158.
111. Hsu, S. K., Lin, L. L., Lo, H. H., and Hsu, W. H. (2004) Mutational analysis of feedback inhibition and catalytic sites of prephenate dehydratase from Corynebacterium glutamicum, *Archives of Microbiology 181*, 237-244.
112. Tan, K., Li, H., Zhang, R., Gu, M., Clancy, S. T., and Joachimiak, A. (2008) Structures of open (R) and close (T) states of prephenate dehydratase (PDT)--implication of allosteric regulation by L-phenylalanine, *Journal of Structural Biology 162*, 94-107.
113. Van Vleet, J., Kleeb, A., Kast, P., Hilvert, D., and Cleland, W. W. (2010) 13C isotope effect on the reaction catalyzed by prephenate dehydratase, *Biochimica et Biophysica Acta 1804*, 752-754.
114. Zamir, L. O., Jensen, R. A., Arison, B., Douglas, A., Albers-Schonberg, G., and Bowen, J. R. (1980) Structure of arogenate (pretyrosine), an amino acid intermediate of aromatic biosynthesis, *Journal of the American Chemical Society 102*, 4499-4504.
115. Byng, G. S., Whitaker, R. J., Shapiro, C. L., and Jensen, R. A. (1981) The aromatic amino acid pathway branches at L-arogenate in Euglena gracilis, *Molecular Cell Biology 1*, 426-438.
116. Hund, H. K., Bär, G., and Lingens, F. (1989) Purification and properties of arogenate dehydrogenase from Actinoplanes missouriensis, *Z Naturforsch [C] 44*, 797-801.
117. Subramaniam, P., Bhatnagar, R., Hooper, A., and Jensen, R. A. (1994) THe dynamic progression of evolved character states for aromatic amino acid biosynthesis in gram-negative bacteria, *Microbiology 140*, 4582-4584.
118. Schenck, C. A., Chen, S., Siehl, D. L., and Maeda, H. A. (2015) Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes, *Nature Chemical Biology 11*, 52-57.
119. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Soll, D., Stetter, K. O., Short, J. M., and Noordewier, M. (2003) The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism, *Proceedings of the National Academy of Sciences of the United States of America 100*, 12984-12988.
120. Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., Rachel, R., and Huber, H. (2008) Nanoarchaeum equitans and Ignicoccus hospitalis: new insights into a unique, intimate association of two archaea, *Journal of Bacteriology 190*, 1743-1750.

121. Podar, M., Anderson, I., Makarova, K. S., Elkins, J. G., Ivanova, N., Wall, M. A., Lykidis, A., Mavromatis, K., Sun, H., Hudson, M. E., Chen, W., Deciu, C., Hutchison, D., Eads, J. R., Anderson, A., Fernandes, F., Szeto, E., Lapidus, A., Kyrpides, N. C., Saier, M. H., Jr., Richardson, P. M., Rachel, R., Huber, H., Eisen, J. A., Koonin, E. V., Keller, M., and Stetter, K. O. (2008) A genomic analysis of the archaeal system Ignicoccus hospitalis-Nanoarchaeum equitans, *Genome Biology 9*, R158.
122. Xue, Y., and Lipscomb, W. N. (1994) The crystallization and preliminary X-ray analysis of allosteric chorismate mutase, *Journal of Molecular Biology 241*, 273-274.
123. Lee, A. Y., Stewart, J. D., Clardy, J., and Ganem, B. (1995) New insight into the catalytic mechanism of chorismate mutases from structural studies, *Chemistry & Biology 2*, 195-203.
124. Fukuchi, S., and Nishikawa, K. (2001) Protein Surface Amino Acid Compositions Distinctively Differ Between Thermophilic and Mesophilic Bacteria, *Journal of Molecular Biology 309*, 835-843.
125. Cambillau, C., and Claverie, J. (2000) Structural and genomic correlates of hyperthermostability, *The Journal of Biological Chemistry 275*, 32383-32386.
126. Berezovsky, I. N., and Shakhnovich, E. I. (2005) Physics and evolution of thermophilic adaptation, *Proceedings of the National Academy of Sciences of the United States of America 102*, 12742-12747.
127. Querol, E., Pérez-Pons, J., and Mozo-Villariá, A. (1996) Analysis of protein conformational characteristics related tothermostability, *Protein Engineering 9*, 265-271.
128. Duzinski, P. K., and Morrison, J. F. (1976) The preparation and purification of sodium prephenate, *Preparative Biochemistry 6*, 113-121.
129. Rieger, C. E., and Turnbull, J. L. (1996) Small scale biosynthesis and purification of gram quantities of chorismic acid, *Preparative Biochemistry & Biotechnology 26*, 67-76.
130. Dawson, R. M. C. (1986) *Data for biochemical research*, Clarendon Press, Oxford.
131. Sambrook, J., Russell, D. (2001) *Molecular Cloning*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring, New York.
132. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680-685.
133. Crowell, A. M. J., Wall, M. J., and Doucette, A. A. (2013) Maximizing recovery of water-soluble proteins through acetone precipitation, *Analytica Chimica Acta 796*, 48-54.
134. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Analytical Biochemistry 72*, 248-254.
135. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Analytical Chemistry 68*, 850-858.
136. Kathiresan, M., Martins, D., and English, A. M. (2014) Respiration triggers heme transfer from cytochrome c peroxidase to catalase in yeast mitochondria, *Proceedings of the National Academy of Sciences of the United States of America 111*, 17468-17473.
137. Laue, T. M., Senear, D. F., Eaton, S. F., and Ross, J. B. A. (1992) Analytical ultracentrifugation in biochemistry and polymer science, (Chemistry, R. S. O., Ed.), pp pp.90-125, Cambridge.
138. Graindorge, M., Giustini, C., Jacomin, A. C., Kraut, A., Curien, G., and Matringe, M. (2010) Identification of a plant gene encoding glutamate/aspartate-prephenate

aminotransferase: the last homeless enzyme of aromatic amino acids biosynthesis, *FEBS Letters 584*, 4357-4360.
139. Kamalampeta, R., and Kothe, U. (2012) Archaeal proteins Nop10 and Gar1 increase the catalytic activity of Cbf5 in pseudouridylating tRNA, *Scientific Reports 2*, 663.
140. Jaworski, A. F., and Aitken, S. M. (2014) Expression and characterization of the Arabidopsis thaliana 11S globulin family, *Biochimica et Biophysica Acta 1844*, 730-735.
141. Leibovitch, M., Bublak, D., Hanic-Joyce, P. J., Tillmann, B., Flinner, N., Amsel, D., Scharf, K. D., Mirus, O., Joyce, P. B., and Schleiff, E. (2013) The folding capacity of the mature domain of the dual-targeted plant tRNA nucleotidyltransferase influences organelle selection, *The Biochemical Journal 453*, 401-412.
142. Mehlin, C., Boni, E., Buckner, F. S., Engel, L., Feist, T., Gelb, M. H., Haji, L., Kim, D., Liu, C., Mueller, N., Myler, P. J., Reddy, J. T., Sampson, J. N., Subramanian, E., Van Voorhis, W. C., Worthey, E., Zucker, F., and Hol, W. G. (2006) Heterologous expression of proteins from Plasmodium falciparum: results from 1000 genes, *Molecular and Biochemical Parasitology 148*, 144-160.
143. Savant-Bhonsale, S., and Cleveland, D. W. (1992) Evidence for instability of mRNAs containing AUUUA motifs mediated through translation-dependent assembly of a > 20S degradation complex, *Genes & Development 6*, 1927-1939.
144. Gaspar, P., Moura, G., Santos, M. A. S., and Oliveira, J. L. (2013) mRNA secondary structure optimization using a correlated stem–loop prediction, *Nucleic Acids Research*.
145. Structural Genomics, C., Architecture et Fonction des Macromolécules, B., Berkeley Structural Genomics, C., China Structural Genomics, C., Integrated Center for, S., Function, I., Israel Structural Proteomics, C., Joint Center for Structural, G., Midwest Center for Structural, G., New York Structural Genomi, X. R. C. f. S. G., Northeast Structural Genomics, C., Oxford Protein Production, F., Protein Sample Production Facility, M. D. C. f. M. M., Initiative, R. S. G. P., and Complexes, S. (2008) Protein production and purification, *Nature Methods 5*, 135-146.
146. Weinglass, A. B., Whitelegge, J. P., Hu, Y., Verner, G. E., Faull, K. F., and Kaback, H. R. (2003) Elucidation of substrate binding interactions in a membrane transport protein by mass spectrometry, *The EMBO Journal 22*, 1467-1477.
147. Geoghegan, K. F., Dixon, H. B., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezzullo, L. H., Martin, L. B., LeMotte, P. K., McColl, A. S., Kamath, A. V., and Stroh, J. G. (1999) Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins, *Analytical Biochemistry 267*, 169-184.
148. Robichon, C., Luo, J., Causey, T. B., Benner, J. S., and Samuelson, J. C. (2011) Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography, *Applied and Environmental Microbiology 77*, 4634-4646.
149. Sampathkumar, P., and Morrison, J. F. (1982) Chorismate mutase-prephenate dehydrogenase from Escherichia coli. Purification and properties of the bifunctional enzyme, *Biochimica et Biophysica Acta 702*, 204-211.
150. Shlaifer, I., and Turnbull, J. L. (2016) Characterization of two key enzymes for aromatic amino acid biosynthesis in symbiotic archaea, *Extremophiles : life under extreme conditions 20*, 503-514.
151. Aponte, R. (2003) Overexpression, purification and preliminary characterization of Prephenate dehydrogenase from the hyperthermophilic bacterium Aquifex aeolicus, Concordia University, Montreal.

152. Khalil, S., Jaworski, I., and Pawelek, P. D. (2014) Identification of a surface glutamine residue (Q64) of Escherichia coli EntA required for interaction with EntE, *Biochemical and Biophysical Research Communications 453*, 625-630.
153. Sterner, R., and Liebl, W. (2001) Thermophilic adaptation of proteins, *Critical Reviews in Biochemistry and Molecular Biology 36*, 39-106.
154. Mayer, F., Kuper, U., Meyer, C., Daxer, S., Muller, V., Rachel, R., and Huber, H. (2012) AMP-forming acetyl coenzyme A synthetase in the outermost membrane of the hyperthermophilic crenarchaeon Ignicoccus hospitalis, *Journal of Bacteriology 194*, 1572-1581.
155. Riepl, R. G., and Glover, G. I. (1979) Regulation and state of aggregation of Bacillus subtilis prephenate dehydratase in the presence of allosteric effectors, *The Journal of Biological Chemistry 254*, 10321-10328.
156. Schmit, J. C., and Zalkin, H. (1971) Chorismate mutase-prephenate dehydratase. Phenylalanine-induced dimerization and its relationship to feedback inhibition, *The Journal of Biological Chemistry 246*, 6002-6010.
157. Segel, I. H. (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*.
158. Schwinck, I., and Adams, E. (1959) Aromatic biosynthesis: XVI. Aromatization of prephenic acid to p-hydroxyphenylpyruvic acid, a step in tyrosine biosynthesis in Escherichia coli, *Biochimica et Biophysica Acta 36*, 102-117.
159. Lutke-Eversloh, T., and Stephanopoulos, G. (2005) Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of Escherichia coli: generation and characterization of tyrosine-insensitive mutants, *Applied and Environmental Microbiology 71*, 7224-7228.
160. Pollastri, G., Baldi, P., Fariselli, P., and Casadio, R. (2002) Prediction of coordination number and relative solvent accessibility in proteins, *Proteins 47*, 142-153.
161. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nature Protocols 10*, 845-858.
162. Xia, T. H., and Jensen, R. A. (1990) A single cyclohexadienyl dehydrogenase specifies the prephenate dehydrogenase and arogenate dehydrogenase components of the dual pathways to L-tyrosine in Pseudomonas aeruginosa, *The Journal of Biological Chemistry 265*, 20033-20036.
163. Schroder, I., Vadas, A., Johnson, E., Lim, S., and Monbouquette, H. G. (2004) A novel archaeal alanine dehydrogenase homologous to ornithine cyclodeaminase and mu-crystallin, *Journal of Bacteriology 186*, 7680-7689.
164. Blumentals, I. I., Robinson, A. S., and Kelly, R. M. (1990) Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaebacterium Pyrococcus furiosus, *Applied and Environmental Microbiology 56*, 1992-1998.
165. Busse, S. C., La Mar, G. N., Yu, L. P., Howard, J. B., Smith, E. T., Zhou, Z. H., and Adams, M. W. (1992) Proton NMR investigation of the oxidized three-iron clusters in the ferredoxins from the hyperthermophilic archae Pyrococcus furiosus and Thermococcus litoralis, *Biochemistry 31*, 11952-11962.
166. Adams, M. W. (1993) Enzymes and proteins from organisms that grow near and above 100 degrees C, *Annual Review of Microbiology 47*, 627-658.
167. Matsui, I., and Harata, K. (2007) Implication for buried polar contacts and ion pairs in hyperthermostable enzymes, *The FEBS Journal 274*, 4012-4022.

168. Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability, *Microbiology and Molecular Biology Reviews : MMBR 65*, 1-43.
169. Ikai, A. (1980) Thermostability and aliphatic index of globular proteins, *Journal of Biochemistry 88*, 1895-1898.
170. Vogt, G., Woell, S., and Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs, *Journal of Molecular Biology 269*, 631-643.
171. Szilagyi, A., and Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey, *Structure Folding and Desig 8*, 483-504.
172. MacBeath, G., Kast, P., and Hilvert, D. (1998) A small, thermostable, and monofunctional chorismate mutase from the archaeon Methanococcus jannaschii, *Biochemistry 37*, 10062-10073.
173. Veronese, F. M., Boccu, E., Schiavon, O., Grandi, C., and Fontana, A. (1984) General stability of thermophilic enzymes: studies on 6-phosphogluconate dehydrogenase from Bacillus stearothermophilus and yeast, *Journal of Applied Biochemistry 6*, 39-47.
174. Coolbear, T., Daniel, R. M., and Morgan, H. W. (1992) The enzymes from extreme thermophiles: bacterial sources, thermostabilities and industrial relevance, *Advances in Biochemical Engineering/Biotechnology 45*, 57-98.
175. Okada, J., Okamoto, T., Mukaiyama, A., Tadokoro, T., You, D. J., Chon, H., Koga, Y., Takano, K., and Kanaya, S. (2010) Evolution and thermodynamics of the slow unfolding of hyperstable monomeric proteins, *BMC Evolutionary Biology 10*, 207.
176. Zuber, H., and Friedman, S. M. (1978) *Biochemistry of Thermophily*, Academic Press, New York.
177. Olszewski, M., Balsewicz, J., Nowak, M., Maciejewska, N., Cyranka-Czaja, A., Zalewska-Piatek, B., Piatek, R., and Kur, J. (2015) Characterization of a single-stranded DNA-binding-like protein from Nanoarchaeum equitans--a nucleic acid binding protein with broad substrate specificity, *PloS one 10*, e0126563.
178. Stetter, K. O. (2006) Hyperthermophiles in the history of life, *Philosophical Transactions of the Royal Society B: Biological Sciences 361*, 1837-1843.
179. Ahmad, S., and Jensen, R. A. (1987) The prephenate dehydrogenase component of the bifunctional T-protein in enteric bacteria can utilize L-arogenate, *FEBS Letters 216*, 133-139.
180. Fischer, R. S., Bonner, C. A., Boone, D. R., and Jensen, R. A. (1993) Clues from a halophilic methanogen about aromatic amino acid biosynthesis in archaebacteria, *Archives of Microbiology 160*, 440-446.
181. Porat, I., Waters, B. W., Teng, Q., and Whitman, W. B. (2004) Two biosynthetic pathways for aromatic amino acids in the archaeon Methanococcus maripaludis, *Journal of Bacteriology 186*, 4940-4950.
182. Xia, T., Song, J., Zhao, G., Aldrich, H., and Jensen, R. A. (1993) The aroQ-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in Erwinia herbicola, *Journal of Bacteriology 175*, 4729-4737.
183. Dopheide, T. A., Crewther, P., and Davidson, B. E. (1972) Chorismate mutase-prephenate dehydratase from Escherichia coli K-12. II. Kinetic properties, *The Journal of Biological Chemistry 247*, 4447-4452.
184. Dudzinski, P. K., and Morrison, J. F. (1976) *Preparative Biochemistry and Biotechnology 6*, 113-121.
185. DeLano, W. L. (2002) The PyMOL Molecular Graphics System

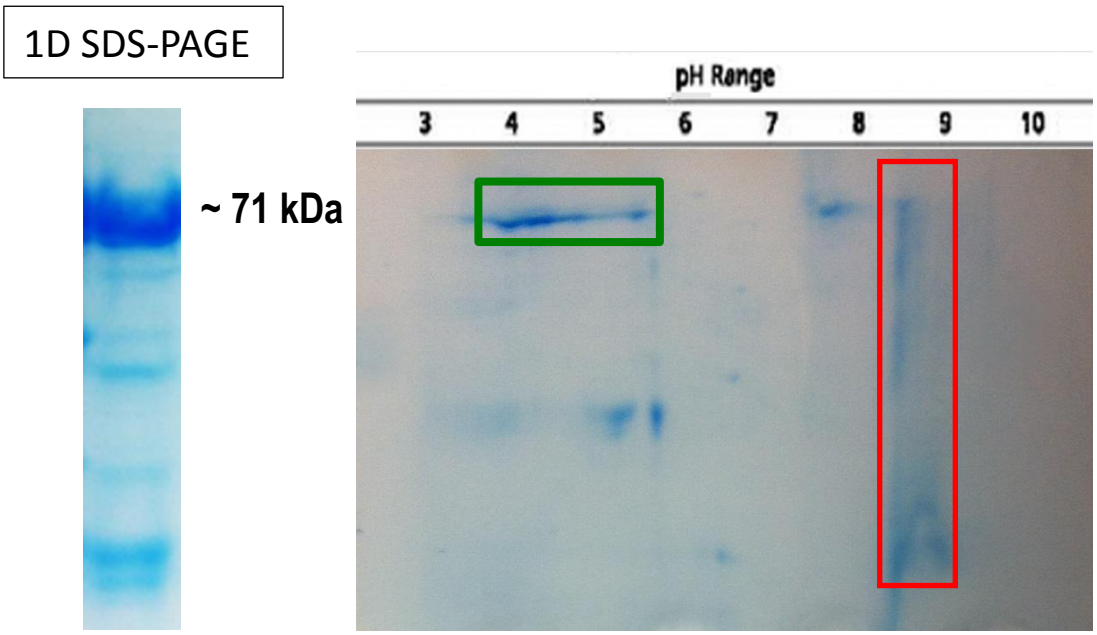
186. Lakowicz, J. (1999) *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers, New York, Boston, Dordrecht, London, Moscow.
187. Schmid, F. X. (1989) *Spectral Methods of Characterizing Protein Conformation and Conformational Changes*, IRL Press, Oxford, UK.
188. Chen, S., Vincent, S., Wilson, D. B., and Ganem, B. (2003) Mapping of chorismate mutase and prephenate dehydrogenase domains in the Escherichia coli T-protein, *European Journal of Biochemistry / FEBS 270*, 757-763.
189. Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability, *Proceedings of the National Academy of Sciences of the United States of America 103*, 5869-5874.
190. Bommarius, A. S., and Riebel-Bommarius, B. R. (2004) *Biocatalysis: Fundamentals and Applications*, Wiley.
191. Zheng, H., Bertwistle, D., Sanders, D. A., and Palmer, D. R. (2013) Converting NAD-specific inositol dehydrogenase to an efficient NADP-selective catalyst, with a surprising twist, *Biochemistry 52*, 5876-5883.
192. Riebel, B. R., Gibbs, P. R., Wellborn, W. B., and Bommarius, A. S. (2003) Cofactor Regeneration of both NAD+ from NADH and NADP+ from NADPH:NADH Oxidase from Lactobacillus sanfranciscensis, *Advanced Synthesis & Catalysis 345*, 707-712.
193. Holmberg, N., Ryde, U., and Bulow, L. (1999) Redesign of the coenzyme specificity in L-lactate dehydrogenase from bacillus stearothermophilus using site-directed mutagenesis and media engineering, *Protein Engineering 12*, 851-856.
194. Rial, D. V., and Ceccarelli, E. A. (2002) Removal of DnaK contamination during fusion protein purifications, *Protein Expression and Purification 25*, 503-507.
195. Benvenuti, M., and Mangani, S. (2007) Crystallization of soluble proteins in vapor diffusion for x-ray crystallography, *Nature Protocols 2*, 1633-1651.

Appendix 1A

gene	Protein name (~ 71 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
NEQ192* of *N. equitans	**PD-CM-PDT**	**Q74NC4**	**148**	**34**	**50**	**71.35**
glmS	**L-gln-D-fru-6-p aminotransferase**	**P17169**	**770**	**30**	**67**	**67**
dnaK	**Chaperone DnaK**	**P0A6Y8**	**169**	**11**	**54**	**69**
groL	**chaperone**	**P0A6F5**	**120**	**2**	**54**	**57**
htpG	**Chaperone Hsp90**	**P0A6Z3**	**71**	**18**	**35**	**71.42**
arnA	**Polymyxin resistance protein**	**P77398**	**579**	**2**	**75**	**75**
rpsA	**30S ribosomal protein**	**P0AG67**	**107**	**22**	**49**	**61**
fusA	**Elongation factor G**	**P0A6M8**	**57**	**11**	**25**	**78**

Orbitrap Velos MS analysis of ~ 71 kDa band digested sample of NeTyrA expressed from cloned genomic DNA and purified by Q-sepharose and Ni-NTA chromatographies. Table lists *N. equitans* PD-CM-PDT and *E. coli* proteins whose peptides were detected by tandem MS analysis.

Appendix 1B



1D and 2D gel analysis of NeTyrA purification by Q-sepharose and Ni-NTA chromatographies. A protein sample of 62 µg was analyzed first by isoelectric focussing followed by SDS-PAGE. Green and red highlight samples which were subjected to in-gel tryptic digestion followed by tandem MS analysis.

Appendix 2A

Position	Calculated Mass	Observed	Peptide Sequence
1-17	1899.9		MGSSHHHHHHSSGLVPR
18-25	858.5	+	GSHMISVK
26-33	888.6	+	ILIIGFGR
34-43	1308.7	+	LGQYFYNYLK
46-50	530.3		GLNVK
51-54	524.3		VYSR
58-66	1124.5	+	EIEENEFSK
69-85	1938.0	+	YAILAIPENSYNEILSK
88-101	1533.8	+	ENNFNGVIIDLASK
103-115	1520.8	+	DVVIPIIEQYGFK
116-133	2111.1		FLSLHPLFGPSIYEEFSK
134-138	571.4	+	IVVIK
139-143	579.3	+	ESTDK
144-163	2470.2	+	SFLQFLDFDLIEMSLEEHEK
164-183	2429.3	+	INELQVVTHLLLISYYHFAR
185-188	504.3		FPIK
189-197	981.5	+	TASAEALYR
198-201	504.3		LSER
202-215	1665.0	+	LLEQNPQILLDIQK
224-228	666.3		ENYIK
232-243	1420.7	+	EVSNNIEDYIPK
246-251	694.4		VEGFSR
252-259	871.6		ALLLNLSK
260-263	560.3		LWNK
268-274	846.5	+	QIMVIDK
275-281	827.6	+	LILDLIK
284-288	560.3		NDLAK
298-304	828.4		NLPIEDK
306-310	704.4		WEIEK
312-319	947.6	+	NILLEFAK
322-340	2259.2		ELNPLYTDQLIELLISWAK
341-365	2703.5	+	HIENPKPWIGVLGPIGSFSDEVALK
375-377	501.2		YYR
379-415	3947.1		ISHIFDALENNEIALGIVPI ENVLGGSVNETLDSLIK
416-419	523.3		YNVK
420-427	986.5	+	IVGEYIHR
428-435	859.5	+	INLCLVGK
443-459	1916.0	+	IISHPQAIAQSMEYISK
461-489	3180.5		FPNAEIVYSNSTSEAISMLDEYSLAITSK
490-498	983.6	+	TAAIFYGLK
499-508	1275.6	+	IYDEHIEDNK
514-519	690.5		FIIIGK
521-526	696.4		QLPNPK
527-542	1753.9	+	YSALVFSLEDKPGSLK
543-550	1014.6	+	EVLEIFHK
551-556	766.4	+	HNINLR
558-564	844.5		LESRPDK
570-582	1617.8	+	YLFYVESDLLNNK
583-588	772.5		ILEEIR
591-595	617.3		ANWVK
596-601	743.4	+	HLGNFR
602-610	1118.6	+	EINEIEFPK

Orbitrap Velos MS analysis of NeTyrA proteolytic fragments generated by trypsin

Appendix 2B

Position	Calculated Mass	Observed 42 kDa band	Observed 30 kDa band	Peptide Sequence
1-13	1428.7	+		GSHMSENPLESLK
18-21	504.3	+		ELDK
25-28	530.3	+		EILR
34-38	635.4			FEIVK
39-44	706.4			EITDTK
46-54	1046.6	+		NLGLPVYDR
57-64	992.5			EEEVMVTR*
65-79	1717.9			TVWGLELGIPQEFTK*
80-83	554.3			EMFK
84-90	849.4	+	+	MILEESK
92-109	991.5	+	+	IQLYTPEK
100-116	1832.9			VYVGIYGYGGMGEQLVK*
117-120	508.3			VFSR
125-130	630.4			VVVTGR
131-134	503.3	+		NLEK
135-140	588.3	+		AEGLAK
144-150	844.4	+	+	VEWGEPK
155-166	1393.8			EVEWLILAVPPK*
167-173	683.4			AVPGLVK
174-180	829.5			ELAPLMR
181-192	1176.6	+	+	SGALLSDISSVK
194-201	930.6	+	+	TLVEEVLK
202-233	3561.9			VLPEYIEYISLHPLFGPEVE PLGETVVVVPVK
234-240	988.5			SYDYWVR
241-263	2648.3	+	+	LVQNIFVSMGFEVITSTPEE HDR
264-284	2251.2	+	+	AMAVTQVLHHFALVSLDEAA K
289-296	1032.4	+	+	EYGVDYMR
297-300	510.3			YATR
305-311	860.5			TLETIQR
314-333	2347.1	+	+	ELSEVIDEIQEMNEYAAHAR
334-338	665.4			EEFLK
339-345	778.4	+	+	VASQMDK

Orbitrap Velos MS analysis of IhTyrA proteolytic fragments generated by trypsin. Shown are peptides with 0 missed cleavages and with a minimum peptide mass of 500 Da.
*Peptides observed in blank (empty gel piece) were excluded from the table.

Appendix 2C

gene	Protein name (~ 40 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
Igni_0892 of *I. hospitalis*	CM-PD	A8AAX2	1978	25	61	41
tufB	Elongation factor	P0CE48	77	7	21	43
gapA	Glyceraldehyde-3-P dehydrogenase	P0A9B2	64	6	26	36
ompF	Outer membrane porin a1	A0A0G3I1B9	56	3	10	39
talB	Transaldolase B	P0A870	37	3	11	35
mdh	Malate dehydrogenase	P61889	22	3	11	32

gene	Protein name (~ 30 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
Igni_0892 of *I. hospitalis*	CM-PD	A8AAX2	405	18	53	41
rplA	50S ribosomal protein L1	P0A7L0	67	5	29	25
gpmA	phosphoglyceromutase	P62707	52	4	17	29

Orbitrap Velos MS analysis of digested protein from ~ 40 kDa and ~ 30 kDa bands of IhTyrA purified by Ni-NTA chromatography and thrombin treated. Table lists *I. hospitalis* CM-PD and *E. coli* proteins whose peptides were detected by tandem MS analysis. SDS-PAGE analysis which separated the two bands is presented in Fig 2.11.

Appendix 3A

Enzyme Activity	Apparrent K_M (mM)		Specific Activity (U/mg)
	Cofactor	Substrate	
CM		**0.7 ± 0.08**	**5 ± 0.3**
PD			
NAD^+	**2.1 ± 0.19**	**0.5 ± 0.09**	**26 ± 0.9**
$NADP^+$	**0.1 ± 0.01**	**0.7 ± 0.14**	**12 ± 0.8**
PDT		**0.2 ± 0.05**	**6 ± 0.4**

Kinetic parameters of the reactions catalyzed by NeTyrA expressed from the cloned gene. Assays were performed as indicated in section 2.1.13. Values were calculated from the initial rates using at least 6 substrate concentrations. CM reaction - chorismate was varied from 0.025 – 1.5 mM. PD reaction – prephenate was varied from 0.025 - 8 mM at a fixed concentration of 40 mM NAD^+ or 2 mM $NADP^+$. When varying the cofactor concentration, $NADP^+$/ NAD^+ (0.05 - 8 mM), prephenate was fixed at 2.5 mM. PDT reaction – prephenate was varied from 0.1 - 2 mM.

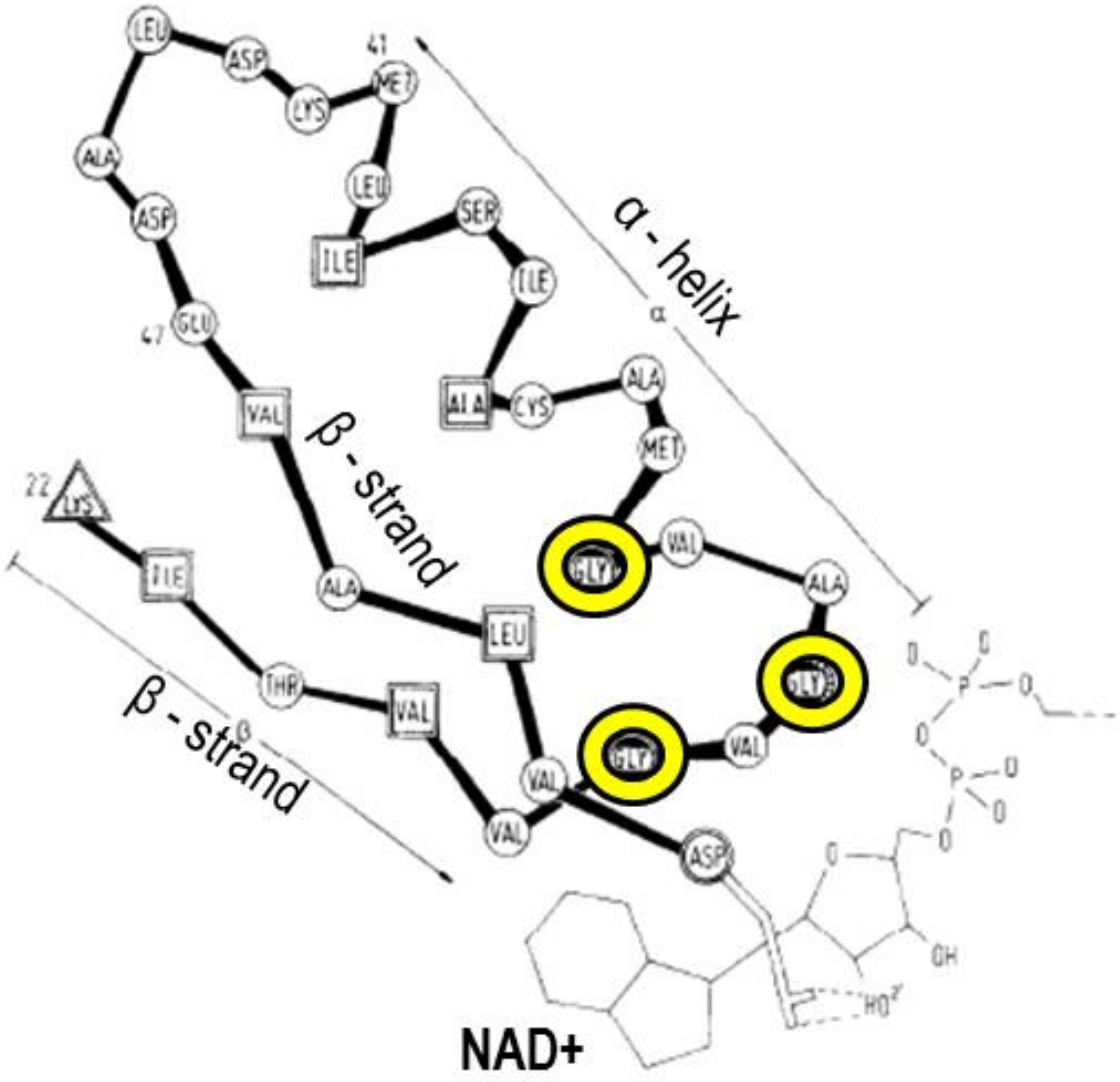
Appendix 3B

gene	Protein name (~ 71 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
NEQ192* of *N. equitans	**PD-CM-PDT**	**Q74NC4**	**432**	**12**	**25**	**71.35**
dnaK	**Chaperone DnaK**	**P0A6Y8**	**158**	**8**	**24**	**69**

gene	Protein name (~ 20 kDa band)	UniProtKB Accession No.	Mascot score	unique peptides	Sequence coverage (%)	MW (kDa)
NEQ192* of *N. equitans	**PD-CM-PDT**	**Q74NC4**	**211**	**14**	**27**	**71.35**
slyD	**Chaperone**	**P0A9K9**	**1776**	**8**	**51**	**21**
rplB	**50S ribosomal protein L2**	**P60422**	**116**	**5**	**25**	**30**
sspB	**ClpXP protease**	**P0AFZ3**	**99**	**5**	**33**	**18**

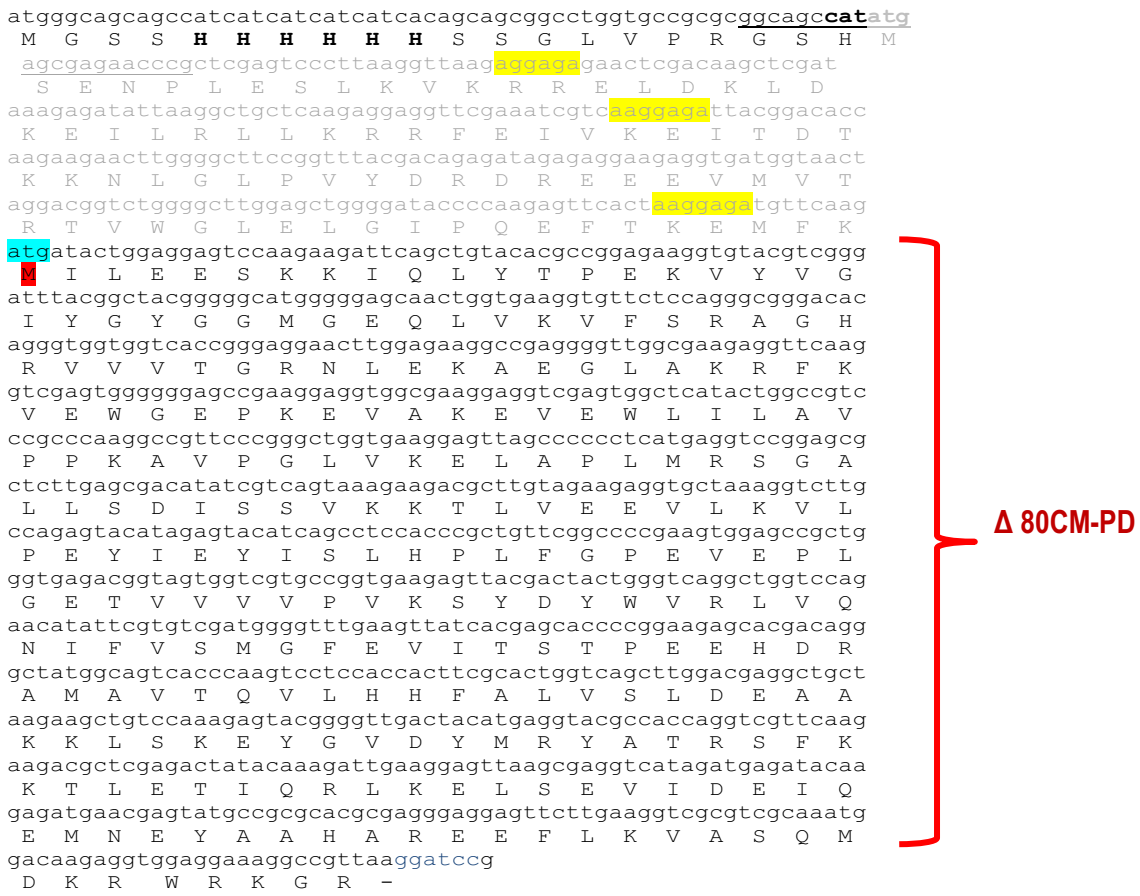
Orbitrap Velos MS analysis of ~ 71 kDa and ~ 20 kDa bands digested sample of NeTyrA heated at 95°C for 3 h. Table lists *N. equitans* PD-CM-PDT and *E. coli* proteins whose peptides were detected by tandem MS analysis. SDS-PAGE analysis of the sample from which the appropriate bands were excised and prepared for MS analysis is shown in Fig 3.8 A

Appendix 3C



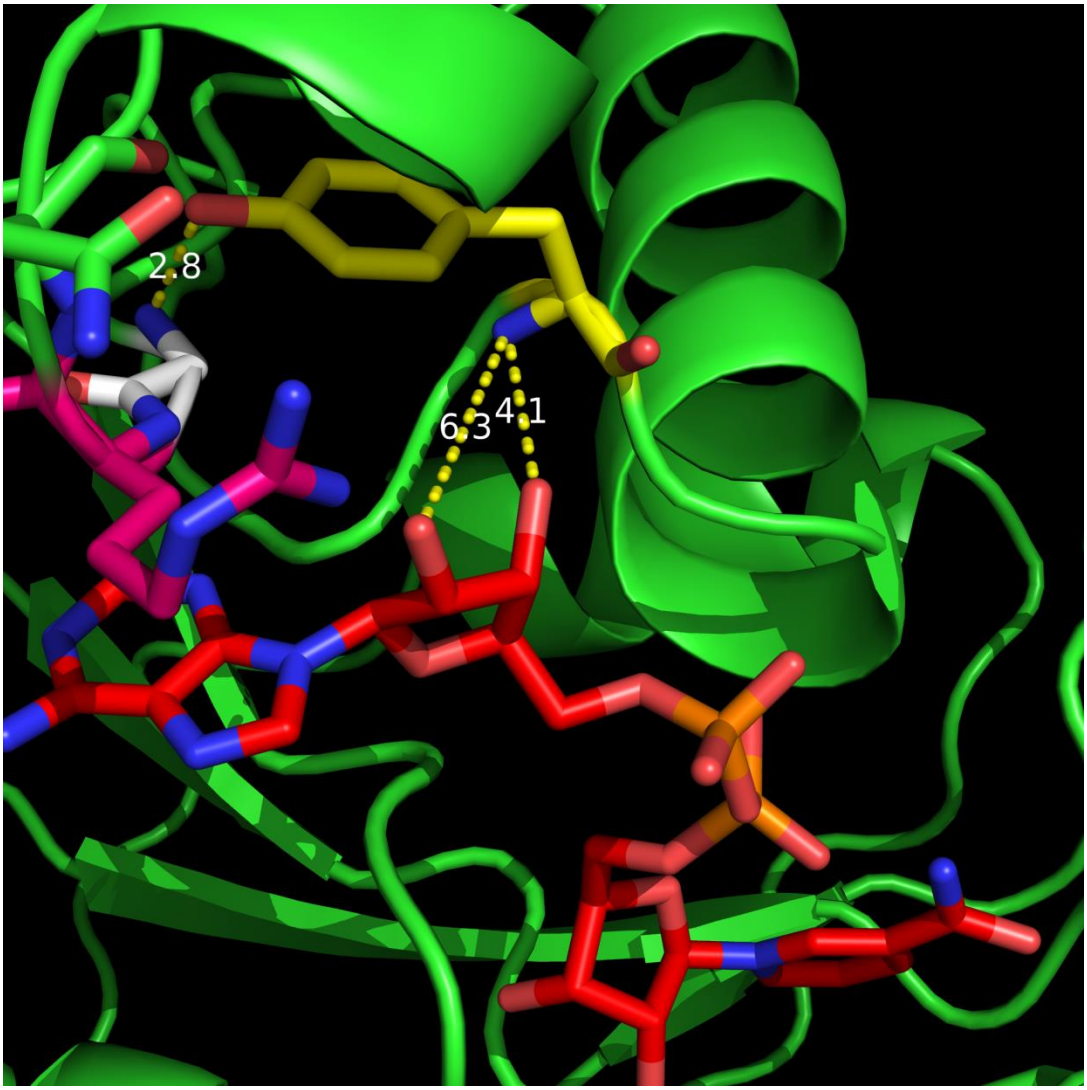
Schematic representation of the ADP-binding βαβ-fold of spiny dogfish M-lactate dehydrogenase. Adapted from [93]. Glycines from the consensus motif GxGxxG are highlighted in yellow. Triangles represent basic or hydrophilic residues, while squares represent small and hydrophobic residues.

Appendix 4A



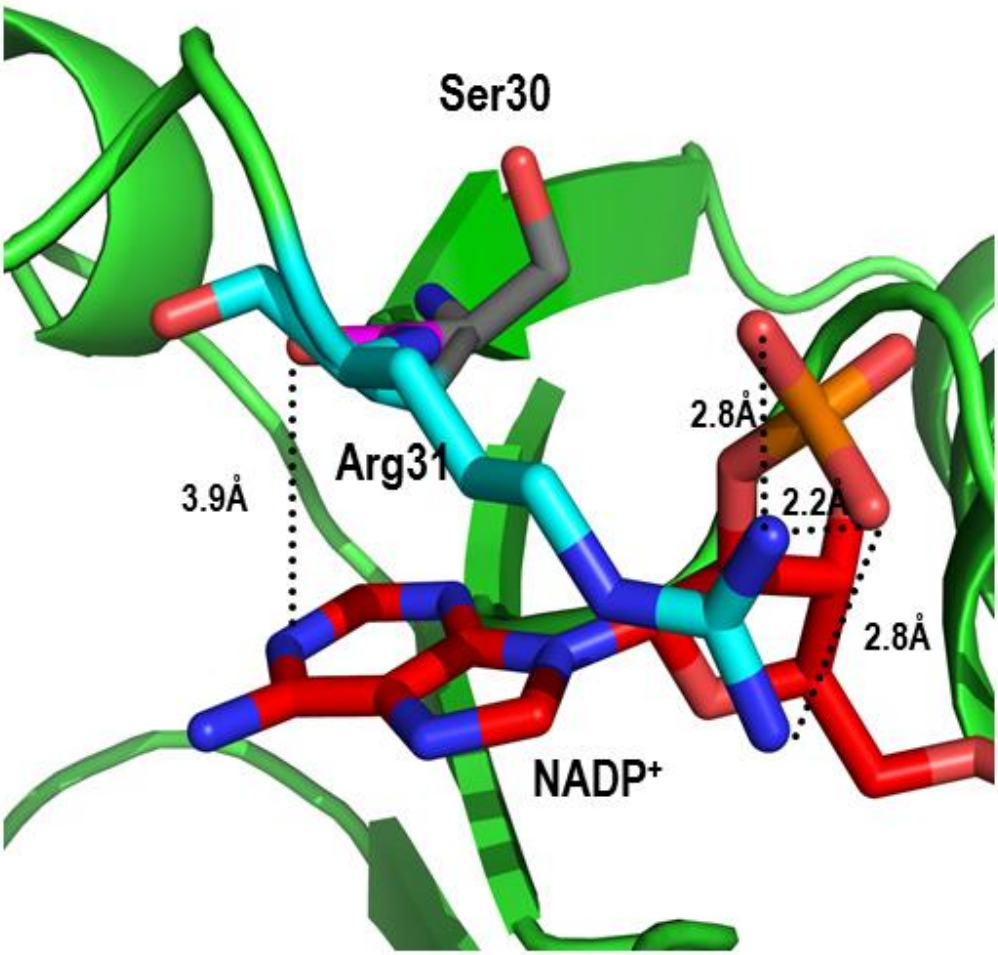
Sequence map of His-tagged IhTyrA. Sequence of the 80 amino acids comprising the CM portion of the protein are highlighted in light gray. The Shine Dalgarno sequence of the ribosome binding site is highlighted in yellow.

Appendix 5A



NAD(P)$^+$ binding site in *I. hospitalis* PD model. Model of *I. hospitalis* PD complexed with NAD$^+$ (red) based on *H. influenzae* PD crystal structure (PDB-2pv7). Gly126, Arg127 and Tyr104 are highlighted in white, pink and yellow, respectively.

Appendix 6A

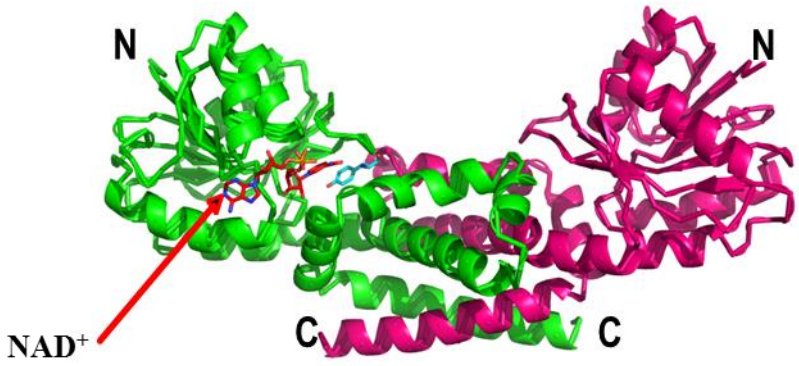


Crystal structure shows interactions of $NADP^+$ interactions with Arg31 in AD from *Synechocystis* sp. PCC 6803. $NADP^+$ is highlighted in red. Ser30 and Arg31 are highlighted in gray and blue, respectively. (PDB-2f1k).

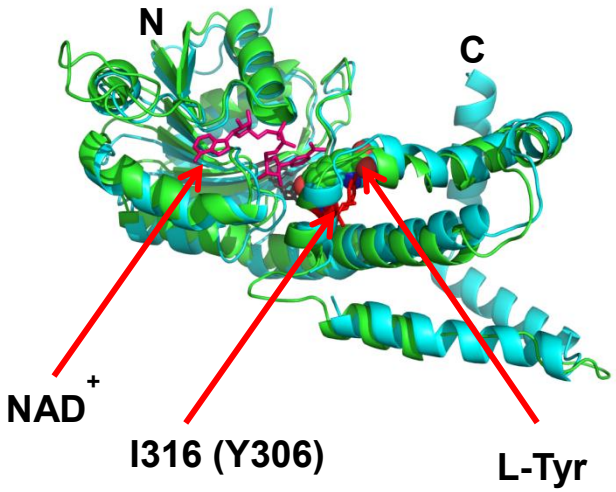
Appendix 7A

Dimeric model of the CM domain of IhTyrA complexed with *endo*-oxabicyclic diacid. The model is based on the crystal structure of CM from *T. thermophilus* PDB-2d8d (~39% sequence identity). The transition state analog, *endo*-oxabicyclic diacid is highlighted in green. Met81 is highlighted in gray.

Appendix 7B

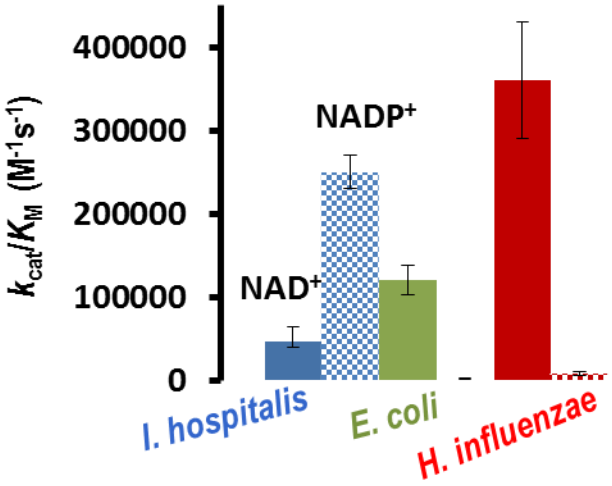


Dimeric model of PD domain of IhTyrA complexed with NAD^+ and L-Tyr. Based on crystal structure of PD from *A. aeolicus* PDB-3ggg, root-mean-square-deviation (rmsd) of 0.77.



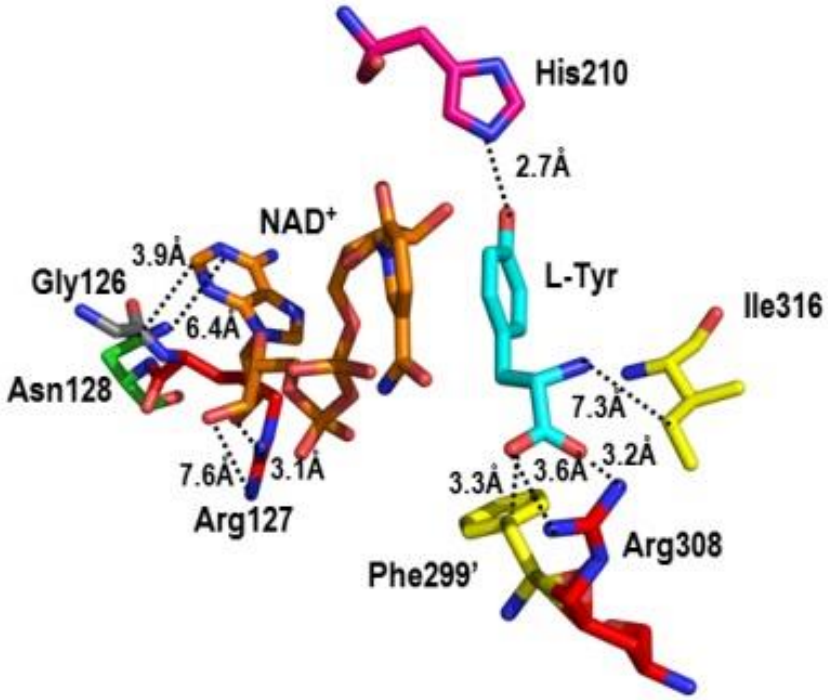
Model of monomeric PD domain of IhTyrA with NAD^+ and L-Tyr overlayed on the crystal structure of PD from *H. influenzae.* The model is based on crystal structure of PD from *H. influenzae* (PDB-2pv7, rmsd-0.53)

Appendix 8A



Cofactor specificity of *I. hospitalis, E. coli and H. influenzae* prephenate dehydrogenases. Filled columns represent values for the efficiency constant k_{cat}/K_M , using NAD^+(filled columns), and $NADP^+$ (checkered columns) as a cofactor in the PD reaction.

Appendix 9A



Active site of modeled dimeric PD domain of IhTyrA in complex with NAD^+ and L-Tyr. The model is based on the crystal structure of PD from *H. influenzae* (PDB-2pv7, rmsd-0.53). The dotted lines indicate the distances between the nearest atoms.