

Multifunctional Enzymes in the shikimate pathway

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Harshul Arora
aus Neu Delhi, Indien

Tübingen
2018

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

06.06.2018

Dekan:

Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter:

Prof. Dr. Andrei Lupas

2. Berichterstatter:

Prof. Dr. Karl Forchhammer

To my mother and father.....

Summary

Multifunctional enzymes are instances of natural fusion proteins, possessing multiple catalytic activities. Via the close proximity of their catalytic domains, they take advantage from concentration of the reactants and may feature allosteric regulation and substrate channeling effects. Substantial progress has been made regarding the kinetic mechanisms of individual enzymes, but the knowledge in the areas of multifunctional enzymes and also transient complexes is rather limited. The seven step shikimate pathway, which is a metabolic route to aromatic compounds, has a number of bifunctional, trifunctional, tetrafunctional and pentafunctional fusion proteins in different organisms. Most notable is the pentafunctional AROM complex, which comprises the central five steps of the pathway, which has withstood structural characterization for decades.

In this work, we shed light on the pentafunctional AROM complex and the three bifunctional enzymes Tm_AroKB, Fp_AroEK and Mb_AroKE of the shikimate pathway by characterizing them functionally and structurally and comparing them to their monofunctional homologs. For the bifunctional proteins, Tm_AroKB and Fp_AroEK the kinetic analysis revealed that their activity is similar to that reported for their monofunctional counterparts. In addition, the kinetic activity of the truncated monofunctional Tm_AroK_{KB} was found to be similar to Tm_AroKB. Interestingly, no allosteric regulation was observed between the two domains in Tm_AroKB. The kinetic analysis of the K domain of Mb_AroKE revealed that it is slower than the other enzymes in this study. However, the genome of the respective organism *Methanoregula boonei* also contains a monofunctional Mb_AroK, which might play the major role in the shikimate pathway.

All crystal structures of the bifunctional enzymes show rigid inter-domain interfaces, with all the catalytic sites solvent accessible. Tm_AroKB is a dimer both in solution and in crystalline state. Remarkably, it contains an intrinsically bound NAD molecule. In all bifunctional proteins, the structure of the individual domains corresponds to that of the monofunctional homologs. However, the domain arrangement in the structures of Fp_AroEK and Mb_AroKE are different. Both arrangements bring the catalytic sites closer and could be part of potential higher order assemblies. Therefore, we hypothesized that the individual enzymes of the pathway form transient complexes for efficient substrate funneling and regulation. *In silico* models of such assemblies were generated based on both Fp_AroEK, and Mb_AroKE, via superimposition on the bifunctional proteins

Tm_AroKB and At_AroDE. These two models are not identical but they both represent a plausible, compact assembly with all active sites solvent accessible.

Finally, we studied the pentafunctional AROM complex. The catalytic turnover of the AROM complex is significantly higher than that of the monofunctional *E.coli* enzymes. The crystal structure of AROM complex, which was found to be different from both *in silico* models, shows that its domains are attached to each other via rigid interfaces and that all the catalytic sites are solvent accessible. Using SAXS and XL-MS, we verified that the conformation of the AROM complex is very similar in solution and in crystalline state. In a computational approach, we further mapped the various conformational states of the individual enzymatic domains from the PDB onto the crystal structure, generating a structural ensemble representing the conformational space of the AROM complex. This ensemble reveals that the complex is optimized for spatial compatibility of the domains, allowing all necessary conformational changes during the catalytic cycle to happen without steric clashes between domains.

Since the shikimate pathway is absent from mammals, it poses a classical drug target. Until now, only glyphosate is utilized as a herbicide for targeting the shikimate pathway. The insight obtained in this study suggests novel approaches to target the shikimate pathway. Most notable is the conformational space of the AROM complex, which is essential in fungi and protists. By targeting the conformational flexibility of the complex, catalytically necessary conformational transitions could be inhibited, which could result in novel, specific fungicides.

Abbreviations

SKM	Shikimic acid
S3P	3-phosphoshikimic acid lithium salt
DHQ	3-Dehydroquinic acid potassium salt
DHS	3-Dehydroshikimic acid
DAHP	3-Deoxy-D-arabinoheptulosonic Acid 7-Phosphate Disodium Salt
NAD	β -Nicotinamide adenine dinucleotide hydrate
NADP	β -Nicotinamide adenine dinucleotide 2'-phosphate disodium salt
NADH	β -Nicotinamide adenine dinucleotide reduced tetrasodium salt hydrate
NADPH	β -Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate
ATP	Adenosine 5'-triphosphate disodium salt hydrate
ADP	Adenosine-5'-diphosphate monosodium salt
PEP	Phospho(enol)pyruvic acid trisodium salt hydrate
GLP	Glyphosate
PK/LDH	Pyruvate Kinase/Lactic Dehydrogenase enzymes from rabbit muscle
EPSP	5-enolpyruvylshikimate-3-phosphate
DNA	Deoxyribonucleic acid
SDS	Sodium dodecyl sulfate
LB	Lysogeny broth
IPTG	Isopropyl β -D-1-thiogalactopyranoside
β -ME	2-Mercaptoethanol
EDTA	Ethylenediaminetetraacetic acid
PNP	Purine ribonucleoside phosphorylase
MESG	2-amino-6-mercapto-7-methylpurine ribonucleoside
CLANS	CLuster ANalysis of Sequences
Coot	Crystallographic Object-Oriented Toolkit
SSM	Secondary-structure-matching
CD	Circular Dichroism
XDS	X-ray Detector Software
GdnHCl	Guanidinium hydrochloride
SAD	Single-wavelength anomalous dispersion
SAXS	Small-angle X-ray scattering
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
TCEP-HCl	Tris (2-carboxyethyl)phosphine hydrochloride
PEG	Polyethylene glycol
EG	Ethylene glycol
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
IMAC	immobilized metal ion affinity chromatography
DTT	Dithiothreitol
PBS	Phosphate buffer saline
NS-TEM	Negative stain Transmission Electron Microscopy
PETRA	Positron-Elektron-Tandem-Ring-Anlage, "positron-electron tandem-ring facility"
SLS	Swiss Light Source
DLS	Diamond Light Source
PSI	Paul Scherrer Institut
PDB	Protein Data Bank
MST	Microscale thermophoresis
XL-MS	Crosslinking and Mass Spectrometry
SC	Screen condition
EM	Electron Microscopy
PCR	Polymerase chain reaction
PMSF	Phenylmethane sulfonyl fluoride

UV	Ultraviolet
EDTA	Ethylenediaminetetraacetic acid
TEV	<i>Tobacco Etch Virus</i>
RMSD	Root-mean-square deviation
CS	Conformational space
LS	Light scattering
EC number	Enzyme Commission number
[S]	Substrate concentration
V_{max}	Maximal velocity
K_M	Michaelis constant
MM	Michaelis-Menten
k_{cat}	Turnover number
[E]	Enzyme concentration
R^2	Coefficient of determination
S.E.	Standard Error
$[E_t]$	Total Enzyme concentration
PB	Protein buffer
CHM	Chorismate
mdeg	millidegrees

Table of contents

Summary.....	3
Abbreviations	5
1. Introduction	11
1.1. Enzymes and Metabolic Pathways.....	11
1.2. Multi-enzyme complexes vs. Multifunctional enzymes	12
1.3. The Shikimate Pathway: Overview	13
1.4. Examination of seven enzymes of the Shikimate Pathway	16
1.4.1. 3-Deoxy-D-arabinoheptulosonate-7-phosphate synthase (DAHPS) (EC 2.5.1.2.54).....	16
1.4.2 3-Dehydroquinate synthase (EC 4.2.3.4).....	17
1.4.3 3-Dehydroquinate dehydratase (EC 4.2.1.10)	18
1.4.4 Shikimate dehydrogenase (EC 1.1.1.25)	20
1.4.5 Shikimate kinase (EC 2.7.1.71).....	21
1.4.6 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (EC 2.5.1.19).....	22
1.4.7 Chorismate synthase (EC 4.2.3.5).....	23
1.5 Multifunctionality in the shikimate pathway	24
1.5.1 Bifunctional dehydroquinate dehydratase-shikimate dehydrogenase (AroDE) .	24
1.5.2 History of AROM complex.....	25
1.6. Rosetta stone approach for predicting protein-protein interactions	27
1.6.1 The Rosetta stone.....	27
1.6.2 Protein-Protein Interactions.....	28
1.6.3. Rosetta-stone proteins in the Shikimate Pathway: Objectives of this study	29
2. Materials and Methods	31
2.1 Materials	31
2.2. Bioinformatic Methods.....	31

2.2.1. Cluster map of the shikimate kinase gene fusions	31
2.2.2. <i>In silico</i> models	32
2.3 Common Biochemical and Biophysical methods	32
2.3.1. Cloning, Protein Expression and Purification	32
2.3.2 Circular Dichroism.....	33
2.3.3 Static Light Scattering	33
2.3.4 Crystallization, crystal freezing, and data collection.....	34
2.3.5 Crystal structure determination	34
2.4 Methods – AroKB.....	35
2.4.1 Cloning, Protein Expression and Purification	35
2.4.2. AroKB kinetics.....	36
2.4.3 Pull down experiments for Tm_AroKB	37
2.4.4 Crystallization and crystal handling.....	38
2.4.5 Denaturation and renaturation	39
2.5 Methods- AroEK.....	40
2.5.1 Cloning, Protein Expression and Purification	40
2.5.2 AroEK kinetics.....	40
2.5.3 Crystallization and crystal handling.....	40
2.6 Methods- AroKE.....	40
2.6.1 Cloning, Protein Expression and Purification	40
2.6.2 Refolding studies of Mp_AroKE	41
2.6.3 AroKE Kinetics	41
2.6.4 Mb_AroKE MST experiments.....	41
2.6.5 Crystallization and crystal handling.....	42
2.7 Methods- AROM complex	43
2.7.1 Cloning, Protein Expression and Purification	43
2.7.2 Kinetic Analysis of the AROM complex.....	43
2.7.3 Crosslinking Mass spectrometry	45
2.7.4 Small Angle X-Ray Scattering (SAXS).....	45
2.7.5 Negative stain Transmission Electron Microscopy screening	46
2.7.6 Crystallization and crystal handling.....	46
2.7.7 Data collection and crystal structure determination AROM complex.....	47

2.7.8. Conformational space Ct_AROM.....	48
2.8. Methods- other proteins	50
2.8.1 Cloning, Protein Expression and Purification	50
2.8.2. Kinetic Analysis of the <i>E. coli</i> enzymes	50
3. Results and Discussion	51
3.1 Cluster analysis of the shikimate kinase fusions	51
3.2. AroKB.....	54
3.2.1 Biochemical and biophysical characterization.....	55
3.2.2 Kinetic analysis of Tm_AroKB.....	57
3.2.2 Crystal structure of Tm_AroKB in complex with SKM	64
3.2.3 Tm_AroKB_SB.....	64
3.2.4 Crystal structure of the Tm_AroKB in complex with S3P and ADP	69
3.3 AroEK.....	70
3.3.1 Biochemical and biophysical characterization.....	70
3.3.2 Kinetic analysis of Fp_AroEK.....	72
3.3.3 Crystal structure of Fp_AroEK in complex with SKM	72
3.4. Mb_AroKE.....	74
3.4.1 Biochemical and biophysical characterization.....	74
3.4.2 Kinetic analysis of Mb_AroKE	76
3.4.2. Crystal structure of Mb_AroKE in complex with SKM and phosphate.....	77
3.6. <i>In silico</i> models of a hypothetical tetrafunctional AROM complex	79
3.7. AROM complex.....	83
3.7.1 Biochemical and biophysical characterization.....	83
3.7.2 Kinetic studies of Ct_AROM.....	85
3.7.3 XL-MS studies Ct_AROM	90
3.7.4. Crystal structure.....	91
3.7.5 Conformational space of Ct_AROM.....	93
3.7.6 SAXS experiments for Ct_AROM	95
3.7.7 Implications of the AROM structure	100
References.....	105

Zusammenfassung.....	119
Contributions	121
Acknowledgements	123
Appendix	125
A.1 Sequences.....	125
A.1.1 <i>Thermotoga maritima</i> shikimate kinase and dehydroquinase synthase (Tm_AroKB).....	125
A.1.2. <i>Thermotoga maritima</i> shikimate kinase cloned in pET His-TEV-Nde-GFP (Tm_AroK _{KB})	126
A.1.3 <i>Thermotoga maritima</i> shikimate kinase and dehydroquinase synthase salt bridge mutant (Tm_AroK _{SB}).....	127
A.1.4 <i>Feacalibacterium prausnitzii</i> shikimate dehydrogenase and kinase (FP_AroEK)	128
A.1.5 <i>Methanoregula boonei</i> shikimate dehydrogenase and kinase (Mb_AroKE) ...	129
A.1.6. <i>Chaetomium thermophilum</i> AROM complex.....	130
A.1.7. <i>Thielavia heterothallica</i> AROM complex cloned in pET28b with NcoI and XhoI	132

1. Introduction

1.1. Enzymes and Metabolic Pathways

The survival of a living organism depends on a network of synchronous interplay of numerous complex chemical reactions. These reactions are needed for all life processes such as energy production, movement, communication, reproduction, excretion and so on [1]. The study of these reactions forms the basis of biochemistry [2]. Together these reactions constitute the metabolism, which comprises the anabolism (synthesis of compounds) and the catabolism (degradation of compounds). However, on their own, most of these reactions are too slow to sustain life. To aid these reactions, complex molecular machines, called “enzymes” come into play. Enzymes are natural catalysts, which greatly accelerate the rates of these chemical reactions with high specificity and play a key role in the existence of life [3]. Enzymology is the study of properties of enzymes.

Probably the first clear record of an enzyme-catalyzed reaction was in 1833 when Payen and Perosz found an alcohol precipitate of a thermolabile substance, which converted “starch to sugar” and named it “diastase” (now called amylase) [4]. In the late 1830s, Louis Pasteur while studying the fermentation of sugar to alcohol by yeast, concluded that fermentation was caused by a vital force contained within the yeast cells called "ferments", which function only within living organisms [5]. In 1878, German physiologist Wilhelm Kühne coined the term enzyme from Greek word “*ενζυμωον*”, (in yeast) to emphasize that there is something in yeast, but not the yeast itself that catalyzes the fermentation reaction [6]. In 1907, Eduard Buchner could isolate a cell-free extract of yeast cells to carry out fermentation. He named this extract, “zymase” and received the first Nobel Prize in enzymology [7]. Since this time, many advancements have been made in the field of enzymology.

Enzymes bind to their substrate molecules with a very high specificity and mediate catalysis to release reaction products [8]. Most of these enzymes are proteins; hence, both the substrate specificity and the rate acceleration are a result of the precise three-dimensional structure of the substrate-binding pockets and the active site [1, 8, 9].

A metabolic pathway thus, can be defined as a sequence of enzyme-catalyzed reactions, which convert one substrate into another [10]. The long list of metabolic pathways includes various pathways involved in carbohydrate metabolism, the amino acid metabolism, the lipid metabolism, and many more [11]. The reactions of the most fundamental metabolic pathways are usually conserved in all organisms, suggesting a common origin and early appearance in the evolution of life [8, 12, 13].

1.2. Multi-enzyme complexes vs. Multifunctional enzymes

In metabolic pathways, the individual enzymes usually perform consecutive catalytic steps and the product of one enzyme acts as the substrate for the next [14]. In some pathways, multiple enzymes form various stable higher order assemblies, to carry out the consecutive steps in a pathway [15].

Multienzyme complexes are groups of noncovalently associated enzymes that catalyze two or more sequential steps in a metabolic pathway linked by common metabolic intermediates [6, 16]. In other words, such multienzyme complexes are massive molecular machines, constituting of different enzymes, which transform the substrate multiple times before releasing the final product. Even prokaryotic cells, which were once thought of as bags of enzymes with random diffusion and no visible compartmentalization, exhibit a high level of coordination of enzymes and are highly efficient in controlling the metabolism in the cytoplasm [17]. The three-dimensional organization can help concentrate reactants to avoid unfavorable reactions, to remove inhibitory products and to channel metabolites from one enzyme to the next [18, 19]. Many instances of multi-enzyme complexes are known, two distinguished examples include the pyruvate dehydrogenase complex (PDH) and the electron transport chain (ETC). PDH is a large multi-enzyme complex with three enzymes and five cofactors, it oxidizes pyruvate to generate acetyl-coA [20, 21]. ETC plays a major role in aerobic respiration, it is a multienzyme complex constituting five different enzymes [14, 22].

On the other hand, multifunctional enzymes are molecules, which contain multiple active sites and catalyze more than one enzymatic reaction. The distinction from multienzyme complexes is that the subunits are not found as discrete enzymes [23]. Multienzyme complexes are products of more than one gene, while a single gene gives rise to a multifunctional enzyme. For example, Tryptophan synthase is a dimeric bifunctional enzyme which shows substrate channeling [24]. Interestingly, fatty acid synthase exists both as a multienzyme complex and as a multifunctional enzyme. The fatty acid synthase (FAS) type 1 in yeast is encoded by two genes, and forms a stable $\alpha_6\beta_6$ dodecameric multienzyme complex [25, 26] while mammalian type-1 fatty acid synthase, encoded by a single gene, forms an α_6 hexamer [26, 27].

When several enzymes compete for the conversion of the same metabolite, the multifunctional enzymes or multienzyme complexes may come to rescue. Some have been shown to transfer the substrate directly from one enzyme to another and to regulate the reaction [24]. These multifunctional enzymes can thus facilitate efficient substrate channeling and/or the metabolic regulation. These enzymes are critical for the communication and cooperation between different functions and pathways within a complex cellular system or between cells [28]. Already the proximity of catalytic sites can facilitate substrate sequestration, as suggested for folate biosynthesis [29, 30].

In the field of multifunctional enzymes, knowledge is still rather limited. The recent genome sequencing projects have helped in adding new candidates to the repertoire. For example, 23 bifunctional enzymes have been reported in plants [29]. Now with the improvement in crystallographic techniques, SAXS and cryo EM, it is a great time to study the complex multifunctional enzymes and multienzyme complexes.

The shikimate pathway is abundant with various multifunctional enzymes with different domain fusions [31-37]. Before discussing the various multifunctional enzymes in the shikimate pathway, I will describe shikimate pathway itself.

1.3. The Shikimate Pathway: Overview

Aromatic amino acids such as Trp, Tyr and Phe, are essential for protein biosynthesis in all living organisms. In addition, derivatives of aromatic compounds such as para-aminobenzoic acid (PABA), vitamin K, and ubiquinone are required for the metabolism of all organisms [38]. The

seven-step shikimate pathway, shown in Figure 1.1, catalyzes the synthesis of a precursor of aromatic compounds in microbes and plants [39, 40].

The first compound to be discovered from this pathway was shikimic acid (shikimate, SKM), from the plant *Illicium religiosum* (aniseed) in 1885. The Japanese name of this plant is shikimi-no-ki, hence the pathway was named the “shikimate pathway” [41, 42]. As shown in Figure 1.1, the precursor of aromatic amino acids, chorismate (CHM) is formed in a seven-step reaction, starting from two carbohydrate precursors, D-erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP). Thereupon, this pathway is a link between the carbohydrate metabolism and the aromatic amino acid metabolism [39]. The synthesis of aromatic amino acids branches at CHM, one branch leading to Trp through anthranilate, while the other branch leads to Phe and Tyr through prephenate (Figure 1.2) [40, 43]. In addition, the aromatic amino acids are precursors of the neurotransmitters serotonin, dopamine, epinephrine, and norepinephrine. Industrially, Phe is used for the production of the artificial low-calorie sweetener aspartame and SKM serves as the starting material for the production of the neuraminidase inhibitor oseltamivir, produced by Roche under commercial name Tamiflu® [44-47]. Some organisms utilize as much as twenty percent of the carbon derived from the carbohydrate catabolism in the shikimate pathway [48].

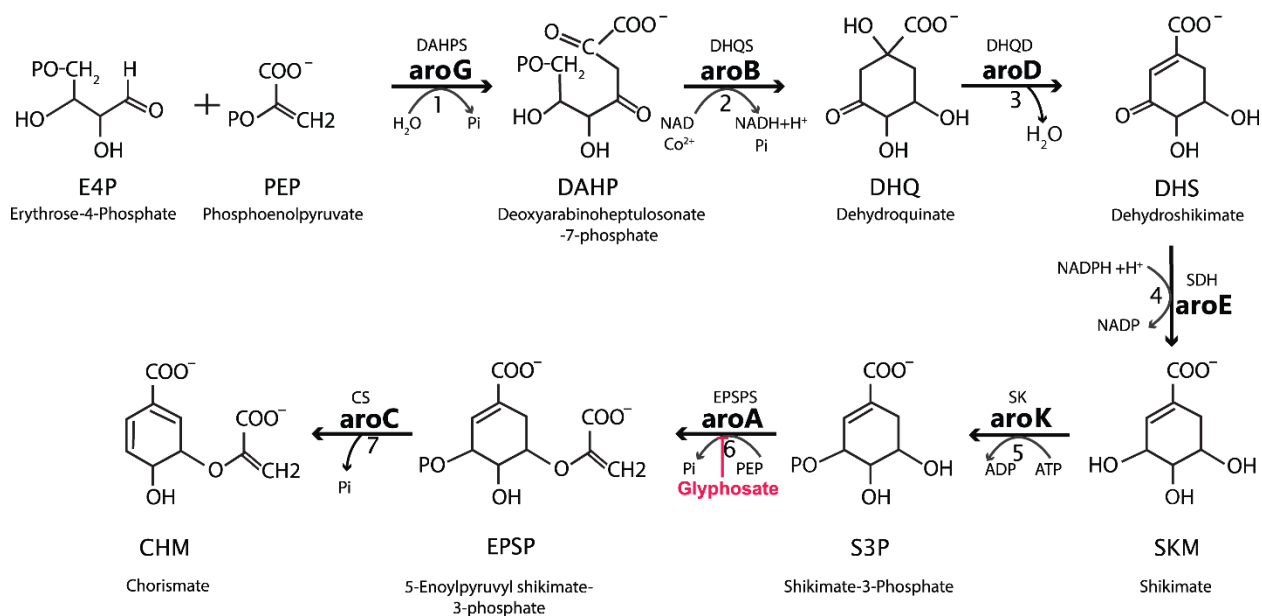


Figure 1.1. The shikimate pathway. The figure shows a schematic representation of seven steps of the shikimate pathway enzymes. The enzymes names are abbreviated and the gene names are shown in bold. Glyphosate, which inhibits the penultimate step in the pathway, is also shown.

The molecular organization and architecture of the shikimate pathway enzymes differs markedly between different taxonomic groups [49], with a differential distribution of the shikimate fusion genes, both in prokaryotes and eukaryotes [35]. With a few exceptions [34, 50-52], the enzymes of the plant shikimate pathway are similar to their prokaryote homologues. Most of the plant enzymes function in plastids, and thus usually contain a plastid transit peptide [32, 33, 35, 39, 53-57].

Animals lack the shikimate pathway enzymes, and must obtain aromatic compounds through diet [38, 39, 43, 58, 59]. Henceforth, the shikimate pathway is a classical drug target [38, 39, 43, 60-63]. Inactivation of the shikimate pathway genes results in a weakened virulence and a decreased survival in a variety of microorganisms [36, 37, 64, 65] including *Mycobacterium tuberculosis* [66] and *Salmonella typhimurium* [67]. Various drug-screening studies have been carried out on *M. tuberculosis* shikimate pathway enzymes [68-75].

5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (AroA) of the shikimate pathway is inhibited by the broad-spectrum, nonselective herbicide glyphosate (GLP) [*N*-(phosphonomethyl)glycine], (Roundup®) [43]. GLP based growth inhibition is a routine test for the presence of a shikimate pathway. For example, the presence of the shikimate pathway was confirmed in the phylum of Apicomplexa, when treatment with GLP could inhibit the growth of *Toxoplasma. gondii*, *Plasmodium falciparum* and *Cryptosporidium parvum* [76].

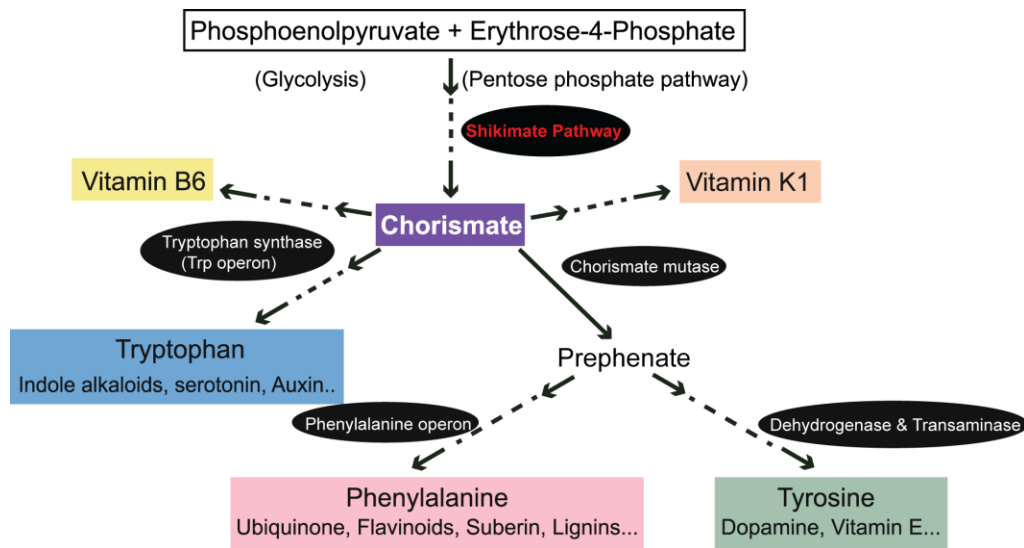


Figure 1.2 Biosynthesis of aromatic compounds from the shikimate pathway: The figure shows the various aromatic compounds that can be synthesized from CHM, which is a product of the shikimate pathway. (Adapted from [32, 36])

1.4. Examination of seven enzymes of the Shikimate Pathway

In most prokaryotes, well-characterized monofunctional enzymes carry out all the steps of the pathway. However, in eukaryotes, several enzymes of the pathway are fused to form various multifunctional assemblies, such as the pentafunctional AROM complex in fungi and Apicomplexans (Section 1.5.1.) [77-79], a bifunctional AroDE in plants [33], and a tetrafunctional AroN in *Acanthamoeba* [37]. Only the crystal structure of the five enzymes constituting the AROM complex are described here.

In the following sub-sections all the enzymes are addressed by their gene names and are called Aro-, as shown in Figure 1.1, while the AROM complex is called as such. The names of substrates are also abbreviated as shown in the Figure 1.1.

1.4.1. 3-Deoxy-D-arabinoheptulosonate-7-phosphate synthase (DAHPS) (EC 2.5.1.2.54)

3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHPS) catalyzes the first committed step of the shikimate pathway, an aldol condensation of phosphoenolpyruvate PEP and E4P to produce DAHP (Figure 1.3) [38, 39, 43, 80]. DAHPSs are metalloenzymes, and require divalent cations such as Mn^{2+} or Fe^{2+} for their activity [43, 81, 82]. The reaction mechanism involves the nucleophilic attack of a water molecule on the C2 position of PEP followed by the addition of C3 of PEP to C1 of arabinose-5-phosphate [38, 83].

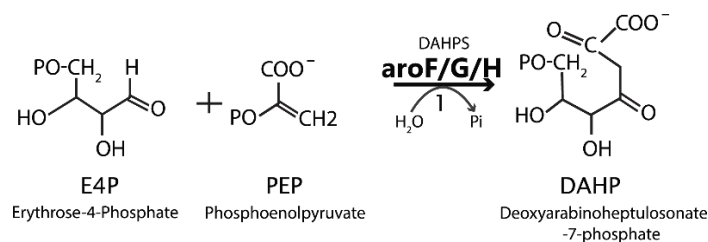


Figure 1.3. DAHPS reaction

DAHPSs can be classified into two unrelated types, type I and type II. The two types share less than 10% sequence identity. The overall fold, the residues that interact with PEP and the divalent metal ion are completely conserved and are almost identically positioned in the two types [84].

Type II enzymes were first observed in plants [85], and later in a few microbial species including *M. tuberculosis* and *Helicobacter pylori*. [84, 86].

Most of the microorganisms with type I DAHPS have three isozymes, which are classified based on their sensitivity to feedback inhibition by one of the three aromatic amino acids. The three isoforms are AroF (Tyr-DAHPS), AroG (Phe-DAHPS), and AroH (Trp-DAHPS) [40, 87]. AroG is the major isozyme found in microorganisms, and constitutes 80% of the total DAHPS activity, while AroF and AroH constitute 20% and 1%, respectively. These enzymes share about 40% of sequence identity with each other [38, 88].

1.4.2 3-Dehydroquinate synthase (EC 4.2.3.4)

Dehydroquinate (DHQ) synthase, (AroB) catalyzes the second enzymatic step of the shikimate pathway. It converts DAHP to DHQ (Figure 1.4) in five steps. The five steps are alcohol oxidation, phosphate elimination, carbonyl reduction, ring opening and intramolecular aldol condensation. They are all catalyzed by a single active site. The enzyme requires NAD and a divalent cation such as Zn^{2+} or Co^{2+} [38, 39, 43, 89, 90] (Figure 1.4). The enzyme avoids undesirable side products by providing a potential conformational template in the final steps [39, 57].

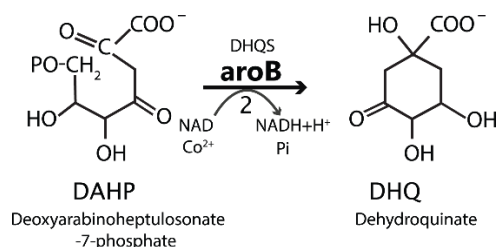


Figure 1.4. AroB reaction

The crystal structure of AroB from *Aspergillus nidulans* (PDB ID: 1DQS) is demonstrated in Figure 1.5. It shows that AroB exists as a functional homodimer containing two AroB subunits. Each monomer of AroB can be divided into two subdomains. First, is the N-terminal α/β domain consisting of a seven-stranded β -sheet, which constitutes the Rossmann fold and serves as NAD-binding site. The orientation of NAD is on the opposite side of the sheet to that seen in all other known Rossmann type NAD-binding proteins. The C-terminal domain is α -helical and contains most of the residues involved in catalysis and in substrate and Zn^{2+} -binding. The mechanism of phosphate elimination requires Arg 130 (colored pink sticks, Figure 1.5) from the other monomer along with other residues [61]. Bacterial AroB is a monofunctional enzyme. The comparison of

crystal structures of bacterial (*S. aureus*) [91] and fungal AroB show that the overall structure of the closed-form is identical in both proteins [92].

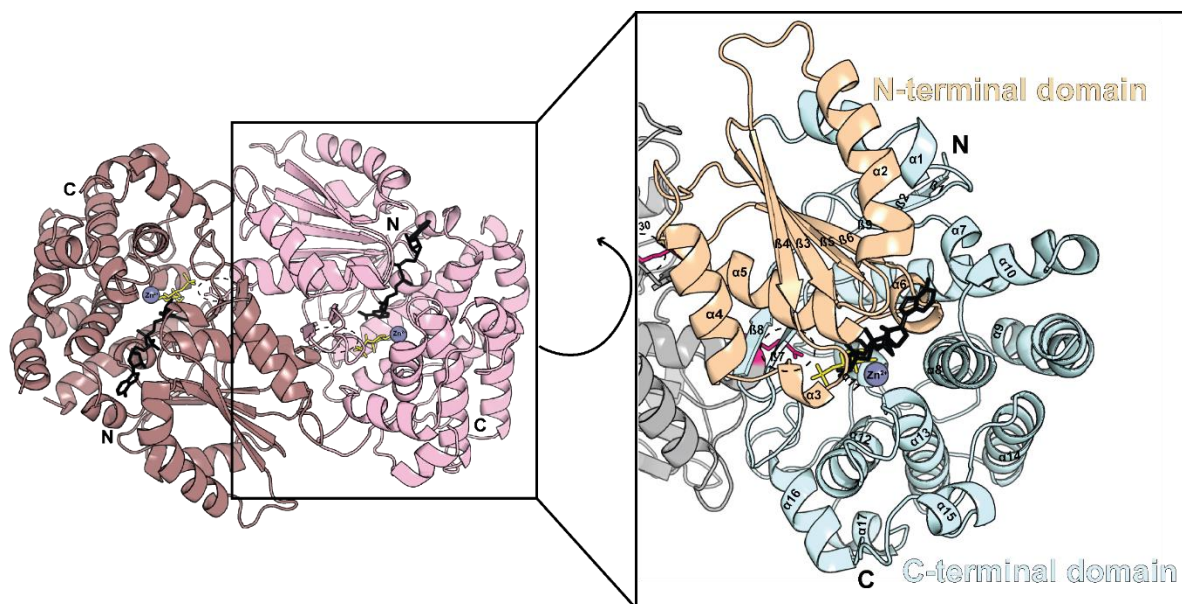


Figure 1.5 Cartoon representation of AroB homodimer (from *Aspergillus nidulans* PDB ID: 1dqs): The figure shows the crystal structure of the AroB homodimer in complex with NAD Zn^{2+} and carbaphosphonate. The complete homodimer is shown on the left, and the two identical monomers are colored in two different colors, pink and light brown. The N and C termini are marked as N and C. One monomer is zoomed on the right and its two constituting domains are highlighted in different colors. The orange color shows the N terminal domain containing the Rossmann fold, which comprises $\beta 3$ - $\beta 6$, and $\alpha 2$ - $\alpha 5$. The Arg 130, which extends contact to the other monomer, is shown as sticks (circled). Zn^{2+} atoms are shown as balls, NAD is shown in black and inhibitor carbaphosphonate is shown in yellow.

1.4.3 3-Dehydroquinate dehydratase (EC 4.2.1.10)

Dehydroquinate dehydratase (AroD) catalyzes the third reaction in the shikimate pathway (Figure 1.6.), which is the dehydration of DHQ to dehydroshikimate (DHS) introducing the first double bond in the ring [43].

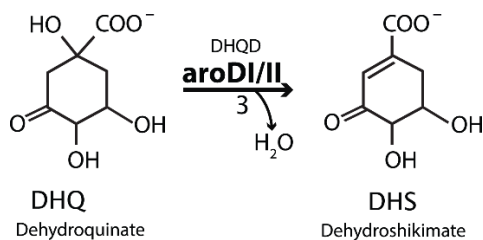


Figure 1.6. AroD reaction

This reaction is common to both the biosynthetic shikimate pathway and the catabolic quinate pathway [40, 93]. AroDs can be divided into two types, AroDI and AroDII [39]. AroDI is primarily found in plants, fungi and many bacterial species, including *E. coli* [94] or *S. typhi* [95, 96], and is exclusively involved in shikimate pathway [94]. In contrast, AroDII can engage in both pathways, and is found in the shikimate pathway of *M. tuberculosis* [96, 97], *S. coelicor* [98] and *H. pylori* [99]. AroDII participates in the quinate pathway in fungi, which utilize quinate as an energy source [60, 100, 101]; it performs a dual function in *A. methanolica*, which can grow on QUN but not on SKM [102]. The two types do not share sequence or structural similarity. AroDI functions as a homo-dimer, and catalyzes a syn-elimination of water using a covalent imine intermediate. AroDII is active as a homo-dodecamer, and catalyzes an anti-elimination of water, with an enolate intermediate. The two enzymes are the result of convergent evolution [103-111].

The crystal structure of AroDI from *Salmonella enterica* is shown in Figure 1.7 (PDB ID: 3M7W) [112]. AroDI is a physiological homodimer containing an eight-stranded α/β -barrel forming a TIM barrel fold [113]. with an additional “N”-terminal anti-parallel β -sheet region, which restricts the substrate entry from one side and finally a flexible lid (LD) that closes on substrate binding [114].

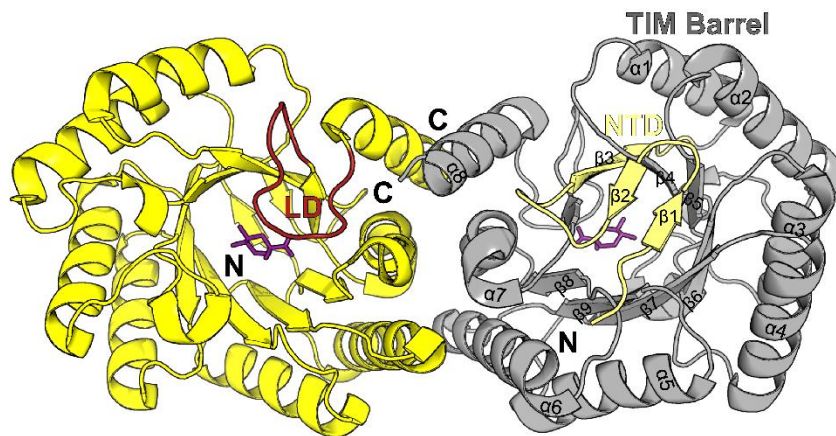


Figure 1.7. Cartoon representation of AroDI (from *Salmonella enterica* PDBID: 3M7W): The figure shows the crystal structure of the complete homodimer of AroDI, the two identical monomers colored yellow and gray. The N and C termini are marked as N and C. One monomer is shown completely in yellow, except for the red lid domain (LD). LD, is in closed conformation as DHQ (pink sticks) is bound. For the second monomer, the N terminal antiparallel β - sheets domain (NTD) is shown in yellow, while the TIM barrel is shown in gray.

1.4.4 Shikimate dehydrogenase (EC 1.1.1.25)

Shikimate dehydrogenase (AroE) catalyzes the fourth step in the shikimate pathway. It catalyzes the reduction of DHS to SKM using NADPH (Figure 1.8). The reaction involves an acid base metabolism [38, 115]. AroE activity is reported in four enzyme classes: AroE, ydiB, SDH-like (SdhL), and AroE-like1 (ael1) [115]. ydiB is a NAD dependent quinate/shikimate dehydrogenase [100, 116]. SdhL from *H. influenzae*, catalyzes the oxidation of SKM with a much lower turnover rate than AroE [117, 118], and ael1 is a recently annotated, functionally distinct SDH subclass that binds SKM with high affinity and exhibits measurable activity with quinate [118].

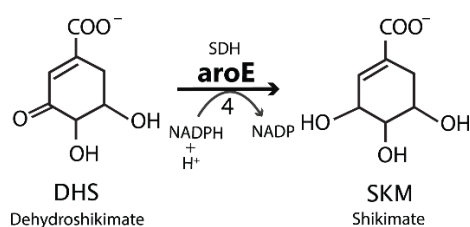


Figure 1.8. AroE reaction

The structure from *Helicobacter pylori* (PDB ID: 3PHI) is illustrated in Figure 1.9, it possesses two structural domains: a catalytic domain and an NADPH binding domain [38]. The domain structure is characteristic of NAD(P)H binding domains in dehydrogenases [119, 120]. Various AroE sequences also contain two strictly conserved *cis*-peptide prolines existing in a sharp turn linking a β -strand and the following α -helix [119].

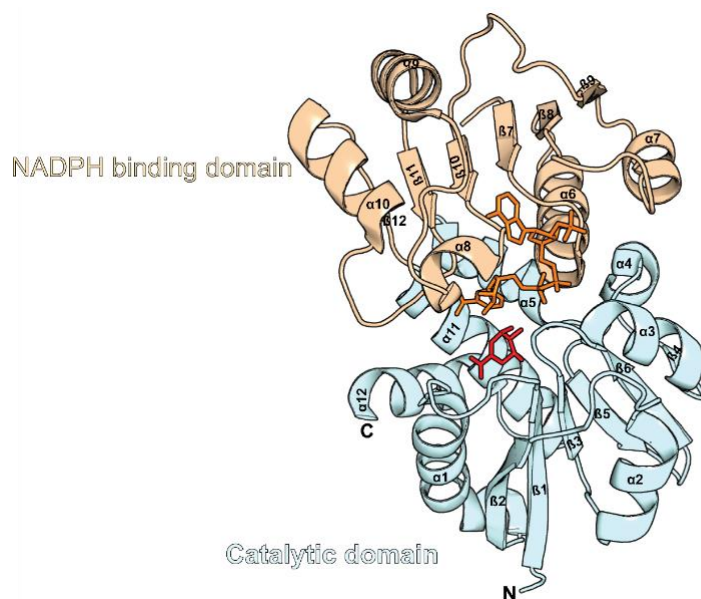


Figure 1.9. Cartoon representation of AroE (from *Helicobacter pylori* PDBID: 3PHI): The figure shows the crystal structure of AroE in two colors. The N and C termini are marked. SKM is shown in red as sticks; NADPH is shown as orange sticks. The catalytic domain is shown in blue, and NADPH binding domain is shown in orange.

1.4.5 Shikimate kinase (EC 2.7.1.71)

Shikimate kinase (AroK/SK) catalyzes the fifth enzymatic step, the phosphorylation of the C₃-OH group of SKM to yield shikimate-3-phosphate (S3P) by utilizing ATP as a co-substrate (Figure 1.10) [38, 39, 44]. AroK requires a divalent cation such as Mg²⁺ for its activity [121].

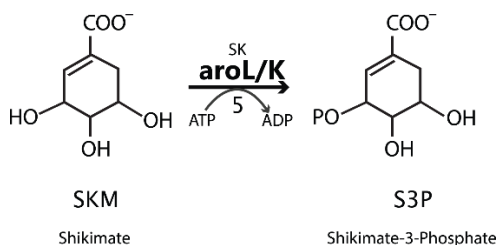


Figure 1.10. AroK reaction

In *E.coli*, SK is present in two isoforms, AroL and AroK, both of which are monomers of 19.5 kDa and contain a type A Walker motif, and share 34% sequence identity [122-125]. Both isoforms catalyze the same reaction *in vitro* but AroL has almost 100 times greater affinity for SKM than AroK [125]. Different plant species have a variable number of isoforms, such as three in *O. sativa* (rice), two in *A. thaliana* and one in tomato [57, 126-128].

The SK crystal structure from *M. tuberculosis* (PDBID: 1U8A) is described in Figure 1.11. AroK belongs to the Nucleoside Monophosphate (NMP) kinase family and constitutes an $\alpha/\beta/\alpha$ fold with a central five stranded parallel β -sheet flanked by eight α -helices. AroKs are subdivided in three domains: Core domain (CD), LID domain (LD) and Nucleotide/SKM binding (SB) domain [129-134]. The core domain includes the central five stranded parallel β -sheet flanked with helices and contains the highly conserved P-loop characteristic of NMP kinase family [135]. The SB domain consists of highly conserved residues involved in SKM binding. The structure comprises three functional motifs, Walker A-motif (P-loop), Walker B-motif and the Adenine binding motif (A-loop) [129-134].

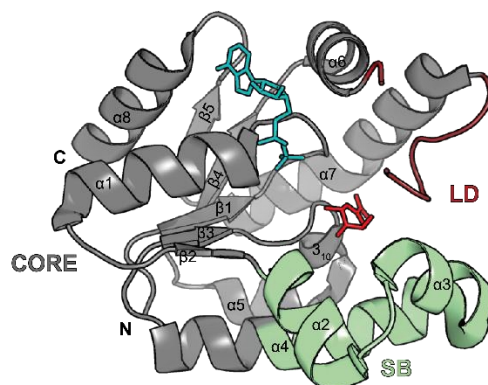


Figure 1.11. Cartoon representation of AroK (from *M. tuberculosis* PDBID: 1U8A): The figure shows the crystal structure of AroK with SKM and ATP bound. N and C termini are marked. The LID (LD) (maroon) connects $\alpha 6$ and $\alpha 7$; the substrate binding (SB) (green) domain consists of $\alpha 2$, $\alpha 3$ and the N-terminal half of $\alpha 4$; and the CORE (gray) constitutes the central five-stranded parallel β -sheet flanked with remaining α helices. SKM is shown as red sticks, while ATP is shown as turquoise sticks.

1.4.6 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (EC 2.5.1.19)

5-enolpyruvylshikimate-3-phosphate (EPSP) synthase AroA catalyzes the sixth and penultimate step of the shikimate pathway (Figure 1.12). In a reversible reaction, it transfers the enolpyruvyl moiety of PEP to the 5-OH position of S3P, and forms EPSP and inorganic phosphate [38, 43]. In the reaction C-O bond cleavage of PEP, exchange of the vinylic protons of PEP occurs [38, 136-138]. This reaction is one out of only four described reactions that catalyze a C-O bond cleavage of PEP instead of a high-energy P-O bond (14.8 kcal/mol).

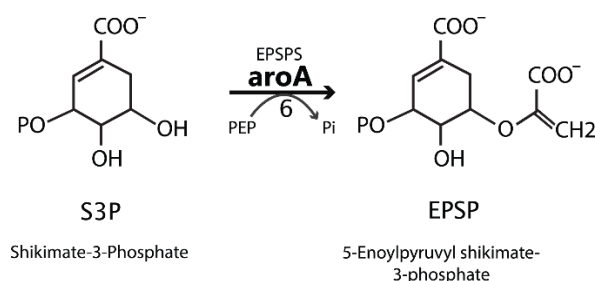


Figure 1.12 AroA reaction

The other three include the drug target Mur A (UDP-*N*-acetylglucosamine enolpyruvyl transferase) which catalyzes the transfer of 3-OH of PEP to UDP-*N*-acetylglucosamine (UDP-NAG) [71]; 3-Deoxy-*D*-manno-octulosonate 8-phosphate (KDO8PS) [139]; and 3-deoxy-*D*-arabinoheptulosonate 7-phosphate (DAHPS) [38].

The nonselective, broad-spectrum herbicide glyphosate [*N*- phosphonomethyl)glycine,] (Roundup[®] Monsanto) acts as a competitive inhibitor of PEP, competing for the same binding site as of PEP, which is very close to S3P [38, 39, 43, 140, 141]. Glyphosate is the only herbicide to target AroA [142]. Based on interaction with glyphosate, AroA enzymes are classified into two classes; glyphosate-sensitive class I AroA and relatively glyphosate resistant class II AroA. Class I includes plants and most bacteria, including *E. coli*; whereas some bacteria, such as *Agrobacterium* sp. strain CP4 form class II, are naturally glyphosate resistant and therefore used to engineer glyphosate resistant crops [43, 141]. Figure 1.12 shows the crystal structure of AroA from *Coxiella burnetii* bound to S3P and Glyphosate (PDB ID: 3SLH) [143].

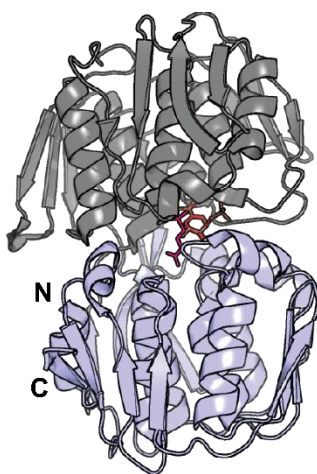


Figure 1.13. Cartoon representation of AroA (from *Coxiella burnetii* PDBID: 3SLH): The figure shows the crystal structure of AroA with SKM, S3P and Glyphosate bound. N and C termini are marked. One domain is shown in gray, while the other one in purple. SKM is shown as red sticks; S3P is shown as brown sticks, and GLP is shown as pink sticks.

The AroA monomer consists of two similar globular domains, an upper and a lower domain, in an open or closed conformation depending on the liganded state. Both the N and C termini of the protein chain are located in the lower domain. Both the domains are remarkably similar, comprising three copies of a $\beta\alpha\beta\alpha\beta\beta$ -folding unit [144]. The active site lies in the cleft between the two domains [38]. Glyphosate binds adjacent to S3P in the PEP-binding site, thereby mimicking an intermediate state of the ternary enzyme–substrates complex [140].

1.4.7 Chorismate synthase (EC 4.2.3.5)

Chorismate synthase (AroC) catalyzes the seventh and final step of the shikimate pathway (Figure 1.14). It catalyzes the trans-1, 4 elimination of phosphate from EPSP (introducing the second double bond in the ring) to yield CHM [38, 39, 43]. The reaction requires reduced flavin (FMN)

as a cofactor for activity even though the overall reaction is redox neutral as AroB [38, 39, 43, 90, 145-150].

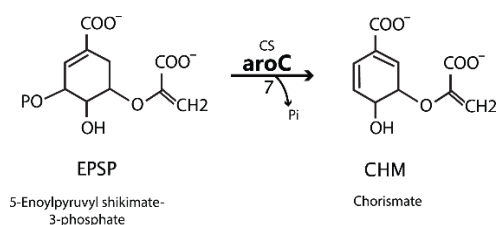


Figure 1.14. AroC reaction

AroCs exist with or without a flavin reductase activity. Monofunctional enzymes, which are usually present in bacteria and plants, do not contain a flavin reductase activity and require reduced FMN. The bifunctional enzymes carry a flavin reductase activity and, therefore, can reduce FMN either using NADPH in fungi or using NADH in *M. tuberculosis* [39, 43, 151]. Monofunctional AroC from *E. coli* is only active anaerobically while the bifunctional enzymes from *N. crassa* are active aerobically [152, 153]. The AroC from *B. subtilis* forms a heterotrimeric complex with a flavin reductase and AroB and is not active in absence of flavin reductase [153, 154].

1.5 Multifunctionality in the shikimate pathway

Various fusion enzymes are seen in the shikimate pathway (section 1.4) but only the bifunctional AroDE found in plants has been characterized structurally and functionally [33]. The crystal structure of AroDE and its possible implications are discussed in the first sub-section. Many attempts have been made to study the architecture and possible role of the pentafunctional AROM complex. Therefore, a short history of the AROM complex is given in the following sub-section. This section ends with a comparison between the different types of fusion enzymes found in the shikimate pathway.

1.5.1 Bifunctional dehydroquinate dehydratase-shikimate dehydrogenase (AroDE)

Bifunctional dehydroquinate dehydratase-shikimate dehydrogenase (AroDE) is the bifunctional enzyme carrying out the third and fourth enzymatic reactions in the shikimate pathway. The reactions include (a) the dehydration of DHQ to DHS to introduce the first double bond in the ring and (b) the reversible reduction of DHS into SKM using NADPH [43].

The substrate DHQ can also be employed as a substrate for the quinate pathway and hence the presence of AroDE can target DHQ to the shikimate pathway [29, 33, 155] [29, 33, 160]. The ratio of turnover rates of E:D have been reported to be around 9:1, thus, DHS is not accumulated but readily converted to SKM, which increases metabolic flux through the shikimate pathway [87, 156].

Singh *et al*, [33] reported the crystal structure of AroDE from *A. thaliana*. (Figure 1.10). The structure contains a type I AroD enzyme, with a missing β hairpin, which caps one opening of the β barrel. They further proposed that the concave architecture of the enzyme would facilitate the transfer of DHS to the E site by increasing the local, effective concentration of DHS at the E site [33] (Figure 1.10).

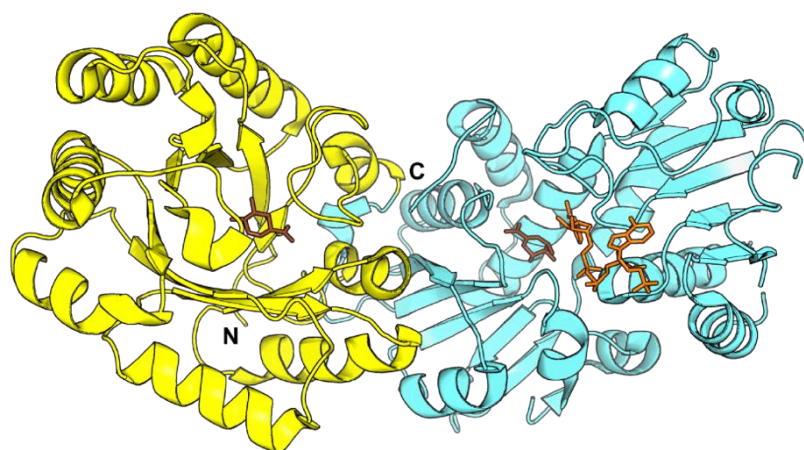


Figure 1.15. Cartoon representation of bifunctional AroDE (PDB ID: 2O7S): The figure shows the crystal structure of bifunctional AroDE bound to DHS and NADPH. The N and C termini are marked as N and C. D domain is shown in yellow, while E domain is shown in blue, DHS is shown in pink and NADPH is shown in orange sticks. Due to the proximity of active sites, this enzyme can facilitate substrate channeling.

1.5.2 History of AROM complex

In eukaryotes, reactions two to six of the shikimate pathway (conversion of DAHP to EPSP), are catalyzed by a single polypeptide called the multifunctional AROM complex [38, 39]. In the 1960s, an *N. crassa* auxotrophic mutant strain Y7655a (aromatic-less: arom), was identified. It could not grow in the presence of SKM, but required phenylalanine, tryptophan, tyrosine and PABA for its growth [157]. In 1967, Giles *et al*, through complementation analysis, identified that this mutant lacks the arom gene cluster. The gene order was BDEKA, and it encoded one protein,

catalyzing five different activities, [158]. Phylogenetic studies suggest that the AROM-gene fusion was an innovation likely to have been present in the progenitor of modern eukaryotes, retained in many lineages, now known to constitute the zygomycete, basidiomycete, and ascomycete fungi, apicomplexa, ciliates, and oomycetes [35, 62].

The early attempts to purify the AROM complex resulted in protein fragments ranging from 165 kDa to 300 kDa [78, 79, 159-162]. The first nucleotide sequence of arom locus from *A. nidulans* showed a single open reading frame and absence of introns [163].

Using glyphosate based growth inhibition; the shikimate pathway was shown to be present in the phylum of Apicomplexans [76]. Apicomplexa are a phylum of diverse obligate intracellular parasites including *Plasmodium* spp., (malaria), *Toxoplasma gondii* (toxoplasmosis) and various opportunistic pathogens of immunocompromised individuals [164]. Bioinformatic analysis has also revealed the presence of the AROM complex in *T. gondii* [36]. Peek *et al*, have characterized the monofunctional E domain of *T. gondii*, where it catalyzes the NADP⁺-dependent oxidation of SKM and is sensitive to AroE inhibitors [165]. Recently, a novel arrangement of gene fusions has been discovered in *A. castellani* AroBAKD, (called AroN) with the E domain fused to two other genes (Figure 1.16) [37].

Despite the discovery in 1967, owing to its large size and labile nature, the structural and kinetic study of the complete AROM complex has been difficult. The few hypotheses for the existence of AROM complex include substrate channeling [164], catalytic facilitation [166], and co-ordinated production of all five enzymes [77, 167].

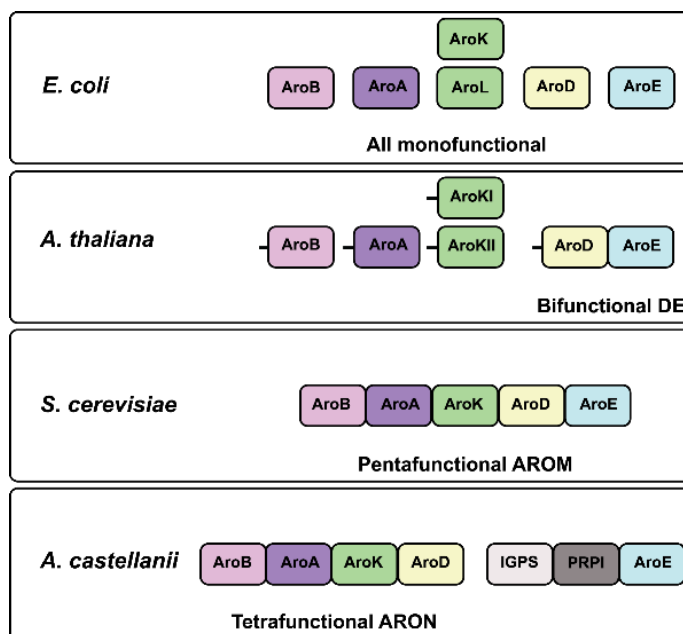


Figure 1.16. Various arrangements of steps two to five in shikimate pathway: The figure shows the various arrangements possible giving rise to different fusion proteins in shikimate pathway. Each taxonomic group is shown with one example organism. The small line before *A. thaliana* enzymes represents the plastid transit signal. IGPS is Indole-3-glycerol-phosphate synthase and PRPI is phosphoribosylanthranilate isomerase (Adapted from [37]).

To summarize the multifunctionality in the shikimate pathway, Figure 1.16 shows the various arrangements of the shikimate pathway enzymes.

1.6. Rosetta stone approach for predicting protein-protein interactions

1.6.1 The Rosetta stone

The Rosetta stone aided in the understanding of Egyptian hieroglyphics (Figure 1.17). It is a black basalt stone, discovered in 1799, near the town of Rosetta. It was engraved during the Ptolemaic Empire at Memphis, Egypt in 196 BC. The carved text contains three different versions of ancient writing, by priests honoring the king of Egypt, Ptolemy V. The top and the middle texts are in Ancient Egyptian using hieroglyphic script and Demotic script, respectively, while the bottom is in Ancient Greek. Remarkably, the Greek passage disclosed that the three scripts had a similar meaning [168]. Thus, it solved the riddle of hieroglyphics, a written language that had been “dead” for thousands of years.

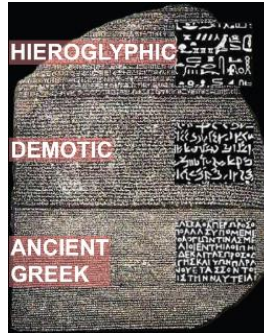


Figure 1.17. The Rosetta stone: The figure shows the Rosetta stone with the same text written on it in three different languages. The first two texts are in Egyptian while the third one is in Greek. (Adapted from [169]).

1.6.2 Protein-Protein Interactions

Marcotte et al, coined the term Rosetta stone method for a computational method, to deduce protein-protein interactions from the genomic sequences based on the observation that some pairs of interacting proteins have homologs in another organism fused into a single protein chain [170]. This method identifies gene-fusion events in complete genomes [171], and predicts functional linkage for protein domains that exist both as fusions and as free polypeptides [172]. It helps in linking distinct proteins together, and suggests the tendency of functional interactions between the linked entities, describing local and global relationships within the proteome [173]. The method is illustrated in Figure 1.18.

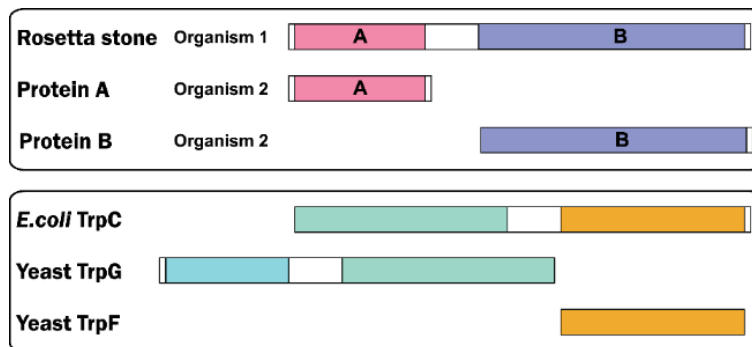


Figure 1.18. The Rosetta stone approach for predicting protein-protein interactions: The top panel shows the principle of the Rosetta stone sequence containing the fusion domains, which are homologous to two separate sequences in other species. The bottom panel shows the example of the fused sequence of TrpC in the *E. coli* genome, which would inform us that the yeast proteins TrpG and TrpF are functionally linked, if it was not known that they both catalyze steps in the biosynthesis of tryptophan. (Adapted from [179])

1.6.3. Rosetta-stone proteins in the Shikimate Pathway: Objectives of this study

Substantial progress has been made regarding the kinetic mechanisms of individual monofunctional enzymes, but our understanding of multifunctional enzymes and interenzyme interactions of transient complexes is rather limited. Owing to their short-lived nature, transient complexes are hard to grasp for structural and functional characterization. Therefore, we exploit “Rosetta-stone” proteins as permanent equivalents of transient complexes. With various multifunctional fusion enzymes present, the shikimate pathway serves as an ideal model system to decipher the potential transient interactions in different organisms. Most remarkable is the pentafunctional AROM complex in fungi, apicomplexans and some protists comprising the central five steps of the pathway. In addition, other multifunctional enzymes include various bifunctional, trifunctional and tetrafunctional fusion enzymes in many organisms.

These different fusions give rise to many questions, such as do the individual enzymes of the shikimate pathway organize to form higher order assemblies during catalysis? If so, do these assemblies vary in different organisms or do they form similar higher order assemblies? To answer these questions we hypothesize that the individual enzymes of the pathway form transient complexes, resembling the AROM complex, for efficient substrate funneling and regulation, during their catalytic cycles. The AROM complex could therefore be the Rosetta-stone of the shikimate pathway.

With this in mind, the following objectives were designed. First, we want to characterize the different bifunctional enzymes and the AROM complex functionally, to test for the existence any inter-domain allosteric regulation. Second, we want to characterize the bifunctional enzymes and the AROM complex structurally, to examine the inter-domain organization. Third, with the redundant sets of crystal structures of several natural bifunctional enzymes we want to examine the Rosetta stone hypothesis and explore possible higher order assemblies *in silico*. Finally, we want to study the pentafunctional AROM complex and compare the structural and functional properties of the AROM complex to the *in silico* assemblies. The gained knowledge will not only help in understanding the shikimate pathway but shall also help to understand how and to which extent the individual enzymes of metabolic pathways are generally organized into transient higher

order assemblies. Furthermore, it may help in the development of drugs targeting protein interactions, and inspire new approaches in metabolic engineering.

2. Materials and Methods

2.1 Materials

The chemicals for preparing the buffers and for protein purification were from the general stock available in the department. The following chemicals were used specifically for this project (Table 2.1)

Table 2.1 List of chemicals used in this project

Name	Product name	Company & lot no.
SKM	Shikimic acid	S5375 Sigma-Aldrich
S3P	3-phosphoshikimic acid lithium salt	55385 Sigma-Aldrich
DHQ	3-Dehydroquinic acid potassium salt	96401 Sigma-Aldrich
DHS	3-Dehydroshikimic acid	05616 Sigma-Aldrich
DAHP	3-Deoxy-D-arabinoheptulosonic Acid 7-Phosphate Disodium Salt	D232050 Toronto Research Chemicals
NAD	β -Nicotinamide adenine dinucleotide hydrate	N7004 Sigma-Aldrich
NADP	β -Nicotinamide adenine dinucleotide 2'-phosphate disodium salt	N0505 Sigma-Aldrich
NADH	β -Nicotinamide adenine dinucleotide reduced tetrasodium salt hydrate	N8129 Sigma-Aldrich
NADPH	β -Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate	N201500 Toronto Research Chemicals
ATP	Adenosine 5'-triphosphate disodium salt hydrate	A26209 Sigma-Aldrich
ADP	Adenosine-5'-diphosphate monosodium salt	02150260 MP Biomedicals
PEP	Phospho(enol)pyruvic acid trisodium salt hydrate	P7002 Sigma-Aldrich
GLP	Glyphosate, PESTANAL®	45521 Sigma-Aldrich
PK/LDH	Pyruvate Kinase/Lactic Dehydrogenase enzymes from rabbit muscle	P0294 Sigma-Aldrich

2.2. Bioinformatic Methods

Pymol was used for visualization and generating pictures [174]. Adobe illustrator was used to design the final pictures.

2.2.1. Cluster map of the shikimate kinase gene fusions

AroK homologs from archaea, bacteria and eukaryotes, were searched in the non-redundant protein sequence database (nr 70), employing two iterations of jackHMMER at default settings [175]. *E. coli* AroK was used as the seed for the first iteration. The resulting sequences were clustered in CLANS to generate a two-dimensional map. CLANS (CLuster ANalysis of Sequences), is a java based application, which adopts the Fruchterman-Reingold graph layout

algorithm to visualize pairwise sequence similarities in either the two-dimensional or the three-dimensional space [176].

2.2.2. *In silico* models

In silico models based on AroEK and AroKE were built using the secondary-structure matching (SSM) algorithm in Coot [177].

2.3 Common Biochemical and Biophysical methods

2.3.1. Cloning, Protein Expression and Purification

The cloning of each gene is described in later sections. All resulting clones were first analyzed with colony PCR, and confirmed by gene sequencing using T7 forward and reverse primers. The BigDye™ Terminator v3.1 Cycle Sequencing Kit was employed for sequencing. Sequencing was carried out in house, with the help of an Applied Biosystems 3730XL DNA Analyzer.

Unless otherwise stated, all the proteins were expressed as C-terminal fusions with a hexa-His tag. BL21(DE3) cells and a temperature of 37 °C were used for protein expression. 5-mL cultures were grown for analytical purposes, while 6-liter cultures were grown for preparative purifications. An overnight starter culture, from a single colony was grown in the lysogeny broth (LB) medium with appropriate antibiotic(s).

The cultures were grown until an OD₆₀₀ of 0.6, and were induced with a final concentration of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expressed for four hours. Expression was analyzed on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). If needed in future, the cell pellet was flash frozen in liquid nitrogen and stored at -80 °C. Protein purification was carried out using immobilized metal ion affinity chromatography (IMAC) [185], followed by gel filtration chromatography [186]; performed on Aekta protein purification systems (GE Healthcare).

The cell pellet was suspended in 50 mL buffer containing 50 mM Tris-HCl pH 8.0, 300 mM NaCl (buffer A), a pinch of lysozyme, a pinch of DNase and a tablet of Roche cOmplete EDTA-free Protease Inhibitor Cocktail. The resulting mixture was lysed by sonication at 50 kHz for four minutes, 3 times with five-minute intervals followed by ultracentrifugation at 126000 X g rpm for 50 minutes to remove the cell debris.

The supernatant was filtered through Millex filters 0.45 μM followed by 0.22 μM (Millex[®] SLHV033RS and SLGP033RS respectively). The filtrate was applied on 20 mL or 5 mL HisTrap HP columns (GE Healthcare).

After loading, the column was washed thrice, first with buffer A, followed by buffer A supplemented with 22.5 mM imidazole and finally with buffer A supplemented with 45 mM imidazole respectively. A gradient elution was made with 45-200 mM imidazole over five column volumes. The fractions were analyzed by SDS-PAGE, pooled and dialyzed overnight in a buffer containing 20 mM Hepes or Tris (pH 8.0), 150 mM NaCl, 2 mM β -mercaptoethanol (β -ME) or Tris (2-carboxyethyl)phosphine hydrochloride (TCEP-HCl) (given separately for individual proteins).

The dialyzed samples were concentrated to a final volume less than 5 mL using Merck 10/30/50/100 kDa Amicon Ultra 15 mL centrifugal filters (depending on protein size). The concentrated sample was loaded onto a HiLoad Superdex 200 26/60 (S 200, 26/60) or HiLoad Superdex 75 26/60 (S 75, 26/60) column (depending on protein size, stated separately for individual proteins) (GE Healthcare).

Gel filtration was performed in the dialysis buffer mentioned above. Fractions containing the active protein were pooled, and concentrated to the required concentrations; aliquots were frozen in liquid nitrogen and stored at $-20\text{ }^{\circ}\text{C}$ until needed.

2.3.2 Circular Dichroism

A JASCO J-810 spectropolarimeter was used to measure the CD spectra from 250-200 nm. Protein concentration was 0.3-0.5 mg/mL in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM β -ME. Thermal denaturation curves over a temperature range of 20-95 $^{\circ}\text{C}$ with a gradient of 1 $^{\circ}\text{C}/\text{min}$ were measured at 220 nm.

2.3.3 Static Light Scattering

Static light-scattering experiments were performed with 1 mg/mL protein (100 μL) in the sample containing buffer (discussed separately for individual proteins) using a GE Healthcare Superdex S75/S200 10/300 GL size-exclusion column connected to a Wyatt MiniDAWN Tristar Laser photometer and a JASCO RI-2031 differential refractometer. Wyatt ASTRA V software was used for data analysis.

2.3.4 Crystallization, crystal freezing, data collection

The proteins were crystallized in the apo state in combination of various ligands. Initial screening was performed in 96-well sitting-drop plates (INTELLI-PLATE® 96 Well, Art Robins Instruments) using the various Qiagen crystal screens and the Morpheus screen (Molecular dimensions). The crystallization screens were set up using the crystallization robots Honeybee 963 (Genomic Solutions) and mosquito® crystal (ttp labtech). The plates were stored at 20 °C and imaged periodically for two months with a Rock Imager 182 (Formulatrix).

The reservoir volume was 50 µl and the drops contained equal volumes of (250 - 550 nl) protein and reservoir solution. If optimization of crystallization conditions was needed, the 96-well sitting drop grid screens were set up using the dragonfly® screen optimization robot (ttp labtech).

The crystals were picked up in a small (<1mm) circular loop. The loops were briefly transferred into a droplet of reservoir solution, which was supplemented with additional cryo-protectant as needed. This loop containing the crystal was flash-cooled and frozen in liquid nitrogen until the measurement. Unless otherwise stated, all the data was collected at the beamline X10SA of the Swiss Light Source (Villigen, Switzerland), using the PILATUS 6M hybrid pixel detector (Dectris Ltd.). The loop containing the crystal was mounted onto the goniometer at the beamline. The data was collected at 100 K and a wavelength of 1 Å.

The detailed crystallization conditions, data collection, and refinement statistics are discussed separately for each protein.

2.3.5 Crystal structure determination

The diffraction data was analyzed, integrated and scaled employing the XDS program suite [178, 179]. The structures of AroKB, AroEK and AroKE were solved using the automated molecular replacement, with the monofunctional homologs, in the program MOLREP (each case discussed individually in later sections) [180]. The structures were refined using Refmac5 [181, 182]. The rebuilding was carried out by using the program Coot [183]. The crystal structure solution of the AROM complex is discussed in section 2.7.7.

2.4 Methods – AroKB

2.4.1 Cloning, Protein Expression and Purification

The genes of AroKBs *Collimonas fungivorans* (Uniprot ID: G0AGG5) (Cf_AroKB), *Acidiphilium multivorum* (Uniprot ID: F0IYK0) (Am_AroKB), *Granulibacter bethesdensis* CGDNIH1 (Uniprot ID: Q0BUI8) (Gb_AroKB), *Oxalobacter formigenes* (Uniprot ID: C3X6K3) (Of_AroKB), were codon optimized for expression in *E.coli* from Eurofins and cloned into the pet22b vector with a C-terminal His tag using NdeI/XhoI restriction sites. Genomic DNA of *Thermotoga maritima* was used as a template to amplify and clone the genes encoding Tm_AroKB (Uniprot ID: Q9WYI3) into pet22b vector with a C-terminal His tag using NdeI/XhoI restriction sites and Tm_AroK_{KB} in pET His-TEV-Nde-GFP (N-terminal hexa-His tag) using NdeI/BamHI. The Tm_AroKB construct was the basis for the construction of the salt bridge mutant, Tm_AroKB_SB (R195A, D315A) (C-terminal His tag), which was built using round the horn mutagenesis [184] (Appendix A.1.1-A.1.13). In addition, the Tm_AroKB construct was the basis for the construction of the FLAG tag containing Tm_AroKB at the N terminal (N-flag-Tm_AroKB). This construct was also cloned with a C-terminal His tag in pet22b.

Cf_AroKB, Am_AroKB, Gb_AroKB, Of_AroKB, Tm_AroKB and Tm_AroKB_SB were expressed in BL21 (DE3) at 37 °C, while Tm_AroK_{KB} was expressed in BL21 (DE3) at 37 °C and BL21 (Gold) at 25 °C. The proteins were dialyzed in the buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl. S 200, 26/60 was employed for the gel filtration of Tm_AroKB and Tm_AroKB_SB; whereas S 75, 26/60 was used for Tm_AroK_{KB}.

Due to the restriction sites, the Tm_AroKB and Tm_AroKB_SB contain the additional sequence Lys, Leu, Ala, Ala, Ala, Leu and Glu prior to the tag. The N-terminal His tag in the Tm_AroK_{KB} was cleaved with TEV protease overnight. Tm_AroK_{KB} and TEV protease were separated with an imidazole gradient; (the protein with no tag precedes the His tag bound TEV protease during elution). After the His tag cleavage, the Tm_AroK_{KB} retained the additional Gly and His prior the first met of the protein sequence.

2.4.2. AroKB kinetics

2.4.2.1 Kinetics for K domain

K domain kinetics was studied for the Tm_AroKB and Tm_AroKB_SB. K domain was kinetically studied by coupling the release of ADP to the pyruvate kinase (PK) (EC 2.7.1.40) and lactate dehydrogenase (LDH) (EC 1.1.1.27) reactions. The SKM/ATP dependent oxidation of NADH was monitored at 340 nm ($\epsilon = 6180 \text{ M}^{-1} \text{ cm}^{-1}$) [185].

The reaction was monitored in a final volume of 0.1 mL in 96 well-plates (Brand plates® 781662) in the Microplate Reader Synergy™ H4 Hybrid or the Synergy™ Mx Monochromator-Based Multi-Mode Microplate Readers at 30°C. The range of substrate (SKM/ATP) from 5.2 μM to 300 μM was analyzed in 11 serially diluted concentrations. The reaction mixture consisted of 50 mM HEPES (pH 8.0), 100 mM KCl, 10 mM MgCl_2 , 0.5 mM ATP, 1.5 mM PEP, 0.3 mM NADH, 9 U of LDH, 6 U of PK, Tm_AroKB, and SKM. While quantifying ATP as a substrate, the range of ATP was varied in the same way, while SKM was kept at a concentration of 0.5 mM (adapted from [186, 187]).

The serial dilutions of substrates were prepared manually in the reaction buffer, in a ratio of 2:3, the highest concentration being 300 μM . Tm_AroKB was added at the end to start the reaction, which was monitored at 340 nm for two hours. The experiments also included a no enzyme blank for each concentration, and a blank without any substrate.

The same reaction was repeated in the presence of the substrates for B domain that is 20 μM of DAHP and NAD. Data analysis was carried out in Sigma plot, discussed in the results section.

2.4.2.2 Kinetics for B domain

B domain kinetics was followed by monitoring the phosphate release employing the, Enzcheck® phosphate release kit [188] and measuring the amount of phosphorylated MESG at 360 nm. The reaction was carried out at 30 °C. The activity was assayed in a final reaction mixture of 0.1 mL in 96-well plates (Brand plates ® 781662).

For estimating DAHP, the reaction mixture contained 0.3 mM NAD, 0.2 mM Co^{2+} , 0.2 mM MESG, 1U of PNPP and 10 nM Tm_AroKB and Tm_AroKB_SB. The range of DAHP from 2 μM to 150 μM with 11 serially diluted concentrations was analyzed. The serial dilutions were prepared manually in the reaction buffer, in a ratio of 2:3, the highest concentration being 150 μM .

The reaction was monitored at 30 °C for two hours. While quantifying NAD as a substrate, the range of NAD was varied in the same way, while DAHP was kept at a concentration of 0.02 mM

The same reaction was repeated in the presence of the substrates for K domain that is 20 μM of SKM and ATP.

2.4.2.3 Data analysis

The Michaelis-Menten model was used for the kinetic analysis. The initial velocity data obtained by varying the concentration of substrate [S] is fitted to the Michaelis-Menten (MM) equation (Eq. 1) to yield maximal velocity values (V_{max}) and the Michaelis constant (K_M). The final Michaelis-Menten (MM) curves were generated in SigmaPlot 12.3 (Systat Software, San Jose, CA, SPSS, Inc) using nonlinear regression analysis with the enzyme kinetics module of SigmaPlot 12.3.

$$v = (V_{max} * S) / (K_M + S) \quad (1)$$

In equation 1, S is the substrate concentration, V_{max} is the maximal rate and K_m is the substrate concentration at which the reaction rate is half its maximal value. Further, the turnover number k_{cat} is the number of substrate molecules converted into product by an enzyme site in a unit time k_{cat}/K_m is the rate constant for the interaction of S and E [189-192]. An error of less than 20% was considered for this study.

2.4.3 Pull down experiments for Tm_AroKB

Tm_AroKB culture was ordered from DSMZ. The cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM KCl, 0.5% Nonidet P-40, 5% glycerol, 1 mM PMSF, protease inhibitor, DNaseI). Samples were sonicated Sonifier (G. Heinemann), and centrifuged at 48,700 x g for 45 minutes at 4 °C to clear the lysates. The supernatant was collected. Meanwhile, anti-FLAG M2 magnetic beads (Sigma) were washed three times with PBS for five minutes. Subsequently N-flag-Tm_AroKB was coupled to the beads by incubating with the protein for 1hr at RT. Equal amounts of cell extract were mixed with coupled beads in binding buffer. 100 μL of the beads sample was used for each sample condition. Samples were incubated for four hours at 4 °C with gentle mixing. After incubation, the resin bead was washed 3 times with 1 mL wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM KCl, 0.5% Nonidet P-40) to remove non-specifically bound proteins. Target molecules were eluted by incubation of the beads with 200 μL of 0.1 M Glycine-HCl, pH 2.0 for 5 minutes with gentle shaking. The resulting sample was neutralized with

1 M Tris and loaded on an SDS gel, the mass spectrometric analysis was performed on a Proxeon Easy-nLC coupled to an LTQ-Orbitrap XL. The data were processed, using software version (1.5.1.0) and had a setting of 1% for the false discovery rate.

2.4.4 Crystallization and crystal handling

Crystallization was carried out both for Tm_AroKB and Tm_AroKB_SB. Table 2.2 lists the final crystallization conditions. To achieve bigger crystals, hanging drops containing 1.5 μ l protein solution and 1.5 μ l of reservoir solution over a reservoir of 500 μ l were set up in EasyXtal plates (Qiagen). To study the activity of Tm_AroKB in crystalline state, the crystals were first soaked in 2.2 M sodium carbonate and later with SKM and ADP for varying time intervals.

Table 2.2 Crystallization and cryo conditions Tm_AroKB

Sr.no.	Protein name	Protein buffer (PB)	Conc. (mg/ml)	Screen condition (SC)	Ligands 10 mM	Cryo
1	Tm_AroKB (sitting drop)	20 mM Tris pH 8.0, 150 mM NaCl	30	(NH ₄) ₂ SO ₄ 2.2 M CH ₃ CO ₂ K 0.2 M	SKM	SC+30% glycerol
2.	Tm_AroKB_SB (sitting drop)	20 mM Tris pH 8.0, 150 mM NaCl	30	(NH ₄) ₂ SO ₄ 1.873 M CH ₃ CO ₂ K 0.1 M	SKM, NAD	SC+30% glycerol

Table 2.3 Data collection and refinement of Tm_AroKB, Tm_AroKB_SB, and Tm_AroKB soaked with SKM and ATP

Data set	Tm_AroKB	Tm_AroKB_SB	Tm_AroKB soaked with SKM and ATP
Wavelength (Å)	1	1	1
Space group	P3 ₁ 2 ₁	P3 ₁ 2 ₁	P3 ₁ 2 ₁
Unit cell (Å)	a = b = 129.53 , c = 87.24, $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$	a = b = 129.41, c = 87.25, $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$	a = b = 130.59, c = 87.83, $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$
A. Diffraction statistics			
Resolution	2.90	2.62	2.750
No. of observations	198638	407829	365220
Completeness	99.8 (99.2)	99.8 (99.0)	99.9 (99.8)
Redundancy	10.45	17.87	14.167
I/sigma(I)	15.14 (1.88)	30.69 (8.05)	34.10 (3.24)
R-meas (%)	11.8 (112.3)	6.1 (24.0)	6.7 (95.9)
CC _{1/2}	99.8 (84.2)	99.9 (98.9)	100 (89.0)
B. Refinement statistics			
Monomers/ASU	1	1	1
Resolution range (Å)	50 - 2.9	50 - 2.62	50 - 2.75
No. of unique reflections	18051	24465	21682
R _{cryst} (%)	23.87	25.53	24.02
R _{free} (%)	28.82	29.37	27.54
Rms bond length deviation	0.0057	0.0052	0.0054
Rms bond angle deviation	1.1810	1.1085	1.1416
Ramachandran statistics (%)	93.87/4.02/2.11	94.93/3.17/1.90	94.29/3.59/2.11

2.4.5 Denaturation and renaturation

Tm_AroKB was denatured in the presence of 6 M guanidinium chloride (GdnHCl). The resulting solution was loaded onto the 20 mL HisTrap HP column, in buffer A supplemented with 6 M GdnHCl. To refold the protein, on column refolding, was attempted using 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 2 M Arg (buffer B) over 30 column volumes. Further, the column was washed with imidazole and Tm_AroKB was purified in the presence of an imidazole gradient (Section 2.1.2.1).

The renaturation was also attempted by diluting the denatured Tm_AroKB ten times in the buffer containing 50 mM Tris (pH 8.0), 500 mM NaCl, 2 mM DTT at 4 °C (buffer C).

2.5 Methods- AroEK

2.5.1 Cloning, Protein Expression and Purification

Eubacterium bifforme genomic DNA was used as a template to clone the gene encoding (Eb_AroEK) (Uniprot ID:B7C8I5) in pET His-TEV-Nco-GFP (N terminal His tag) using the restriction sites NcoI/XhoI. The AroEK gene from *Faecalibacterium prausnitzii* (Fp_AroEK) (Uniprot ID: C7H179) was codon optimized for expression in *E. coli* (Eurofins) and cloned in pet22b (C-terminal His tag) using NdeI/XhoI restriction sites (Appendix A.1.4). Owing to the restriction sites, the Fp_AroEK sequence contained additional Leu and Glu before the His tag.

Eb_AroEK was expressed in BL21 (DE3) at 37 °C. Fp_AroEK was expressed in BL21 (DE3) at 37 °C and in BL21 (pRare) at 25 °C. It was dialyzed in the buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl, 2 mM β -ME. Gel filtration was carried out using S 75, 26/60.

2.5.2 AroEK kinetics

Only the K domain kinetics for ATP was studied for Fp_AroEK in the same way as described for Tm_AroKB (section 2.4.2).

2.5.3 Crystallization and crystal handling

Crystals for Fp_aroEK of sufficient quality were obtained directly from the initial screen. Table 2.4 lists the final crystallization conditions for Fp_AroEK. The data collection and refinement statistics are discussed in Table 2.6 together with Mb_AroKE.

Table 2.4 Crystallization conditions of Fp_AroEK

Protein	Protein buffer	Conc. (mg/ml)	Screen condition (SC)	Ligands 10 mM	Cryo
Fp_AroEK (sitting drop)	20 mM Tris pH 8.0, 150 mM NaCl, 2 mM β -ME	20	Morpheus D10 0.12 M Alcohols, 0.1 M Buffer 3 (pH 8.5) 50% P3	SKM	SC+20% PEG 200

2.6 Methods- AroKE

2.6.1 Cloning, Protein Expression and Purification

Methanoplanus petrolearius genomic DNA was used as a template to clone the gene encoding Mp_AroKE (Uniprot ID: E1RGC9) in pET His-TEV-Nco-GFP (N terminal His tag) using the restriction sites NcoI/XhoI (Appendix 1.7.).

The AroKE gene from *Methanoregula boonei* Mb_AroKE (Uniprot ID: A7I8L9) was codon optimized for expression in *E. coli* (Eurofins) and cloned in pet22b (C-terminal His tag) using NdeI/XhoI restriction sites (Appendix 1.8). Owing to the restriction sites, the sequence contained the additional Leu and Glu before the His tag.

The Mp_AroKE gene was expressed in BL21(DE3) cells both at 37 °C and 25 °C. Mb_AroKE was expressed in BL21 (DE3) at 37 °C and ArcticExpress cells at 10-12 °C. Mb_AroKE was dialyzed in the buffer containing 20 mM Tris (pH 8.0), 150 mM NaCl, 2 β-ME. Gel filtration of Mb_AroKE was carried out using S75, 26/60.

2.6.2 Refolding studies of Mp_AroKE

Mp_AroKE was exclusively found in the inclusion bodies. To solubilize the inclusion bodies, buffer A supplemented with different denaturants such as either 6 M GdnHCl or 8 M urea was added to the pellet obtained after the sonication and ultracentrifugation step. After solubilizing and ultra-centrifuging, the supernatant was loaded onto the 20 mL HisTrap HP column, and denatured Mp_AroKE was eluted with an imidazole gradient, in the presence of denaturants.

In an attempt to refold, Mp_AroKE was serially dialyzed against the buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, 2 mM DTT, GdnHCl / urea and 2 M Arg at 4 °C, while the concentrations of the denaturants and Arg were decreased at each step.

Another attempt to refold the protein was carried out by diluting the denatured Mp_AroKE ten times in buffer C at 4 °C.

2.6.3 AroKE Kinetics

In addition to the substrate range discussed in section 2.4.2, a range of 11 ATP concentrations from 52 μM and 1.5 mM was also tested. The dilution series were prepared manually with a factor of 2:3 with 1.5 mM being the highest concentration.

2.6.4 Mb_AroKE MST experiments

A serial 1:1 dilution of SKM ranging from 122 nM to 2 mM was prepared and mixed with 1 mM Mb_AroKE. MST measurements were performed with a Monolith NT.LabelFree (Nanotemper), using various MST power and laser intensity settings to test the general validity of the obtained data at a temperature of 25°C, using MST power 20% and laser intensity 20 and 40%. The data was analyzed using the NT Analysis 1.5.41 software (Nanotemper).

2.6.5 Crystallization and crystal handling

Initial screens of Mb_AroKE gave sufficient quality crystals in complex with SKM Table 2.4 lists the crystallization conditions for Mb_AroKE while Table 2.5 lists the data collection and refinement statistics.

Table 2.5 Crystallization conditions of Mb_AroKE

Protein	Protein buffer	Conc. (mg/ml)	Screen condition (SC)	Ligands 10 mM	Cryo
Mb_AroKE (sitting drop)	20 mM Tris pH 8.0, 150 mM NaCl, 2 mM β -ME	41	JCSG 48 0.04 M K_3PO_4 , 16% PEG 8000, 20% glycerol	SKM, ADP	SC+10% glycerol

Table 2.6 Data collection and refinement of Fp_AroEK and Mb_AroKE

Data set	Fp_AroEK	Mb_AroKE
Wavelength (Å)	1	1
Space group	P4 ₂ 2 ₁ 2	P2 ₁
Unit cell (Å)	a = b = 154.5, c = 56.7, $\alpha = \beta = \gamma = 90.0^\circ$	a = 43.72, b = 123.80, c = 46.65, $\alpha = \gamma = 90.0^\circ$, $\beta = 97.76^\circ$
A. Diffraction statistics		
Resolution	2.70	2.15
No. of observations	510440	181885
Completeness	99.8(99.3)	98.6 (93.9)
Redundancy	26.14	3.47
I/sigma(I)	19.71 (2.66)	9.21(1.87)
R-meas (%)	14.9 (142.8)	12.5(75.8)
CC _{1/2}	99.9 (81.9)	99.4 (68.5)
B. Refinement statistics		
Monomers/ASU	1	1
Resolution range (Å)	43.94-2.70	35.48 – 1.99
No. of unique reflections	18513	31476
R _{cryst} (%)	23.70	18.0
R _{free} (%)	26.26	22.76
Rms bond length deviation	0.0043	0.0183
Rms bond angle deviation	0.8936	1.8902
Ramachandran statistics (%)	97.53/2.22/0.25	96.85/3.15/0

2.7 Methods- AROM complex

2.7.1 Cloning, Protein Expression and Purification

The codon optimized AROM genes, (~ 4.7 kb) from *Chaetomium thermophilum* (Uniprot ID: G0S061) Ct_AROM and *Thielavia heterothallica* Th_AROM (Uniprot ID: G2QMW9) were obtained as two overlapping strings (~ 2.3 kb each) from Thermofisher scientific. The strings overlapped in 27 and 26 base pairs for Ct_AROM and Th_AROM respectively. The two strings were combined using the overlapping PCR (Appendix 1.7 and 1.8).

The resulting full-length genes were cloned in pET 28b using the restriction sites NcoI and XhoI. Due to the size of the AROM complex, the sequence was validated by sequencing with various primers within the gene. Owing to the vector, the sequence of the AROM complex contains additional Leu and Glu prior to the His tag. The AROM complex was dialyzed in the buffer containing 20 mM Hepes pH 8.0, 150 mM NaCl, 2 mM TCEP. The AROM complex was concentrated using a dialysis bag of 6 kDa MWCO, in the presence of PEG 20000 (no gel filtration was carried out).

For expressing the selenomethionine labeled Ct_AROM complex, the overnight starter culture was grown in LB medium. The following day the cultures were centrifuged, and the cells were washed twice with phosphate buffered saline (PBS) to remove residual LB. The cells were resuspended in minimal media and divided equally in 6 L of minimal media. The cultures were grown at 30 °C until the OD of 0.6, and induced with 1 mM IPTG. The protein was purified in the same way as the native Ct_AROM complex.

Per liter of M-M contained M95X [193] (200 mL), 1 M MgSO₄ (2 mL), 1 M CaCl₂ (0.1 ml), 20% glucose solution (20 ml), 1% thiamine (0.2 mL), Lys, Thr, Phe, Leu, Ile und Val (ca. 50 – 60 mg each) DL-Selenomethionine (ca. 50 – 60 mg).

2.7.2 Kinetic Analysis of the AROM complex

2.7.2.1 Assay for BDE domains

BDE domains were assayed together, measuring the consumption of NADPH at 340 nm ($\epsilon = 6180 \text{ M}^{-1} \text{ cm}^{-1}$) [194] (Figure 2.2). The activity was assayed in 384-well plates, (Thermoscientific 242757); the final volume of the assay mixture was 25 μl . The DAHP range of 0.02 to 0.6 mM with 15 serially diluted concentrations was analyzed. The reaction mixture contained 50 mM

HEPES (pH 8.0), 100 mM KCl, 10 mM MgCl₂, 0.2 mM NAD, 0.2 mM CoCl₂, 0.8 mM NADPH, 0.8 mM ATP, 0.8 mM PEP, 10 nM AROM. The serial dilutions of DAHP in a ratio of 2:3, were made in the buffer containing all the cofactors, employing the Piro® pipetting robot, (Dornier Labtech systems GmbH).

The reaction was carried out with a temperature range from 30 to 50 °C. The plate was covered with the lid and was incubated at the required temperature for 15 minutes on a heating block. After the incubation, the enzyme was added and the reaction was followed for 2 hours at 340 nm. The AroBDE reaction was also performed in absence of ATP and PEP. For studying the effect of GLP on BDE, the BDE assay was performed in the presence of 0.2 mM GLP.

2.7.2.2 Assay for KA domains

In addition, the reactions of KA domains were followed by monitoring the phosphate release employing the Enzcheck® phosphate release kit [188] and measuring the amount of phosphorylated MESG at 360 nm. The reaction was carried out at 30 °C, as MESG is not thermostable. The activity was assayed in a final reaction mixture of 100 µL in 96-well plates (Brand plates ® 781662).

The reaction mixture contained 0.2 mM ATP, 0.2 mM PEP, 0.2 mM MESG, 1U of PNPP and 10 nM AROM. The range of SKM from 2 µM to 150 µM with 11 serially diluted concentrations was analyzed. The serial dilutions were prepared manually in the reaction buffer, in a ratio of 2:3, the highest concentration being 150 µM. The reaction was monitored at 30 °C for two hours.

For measuring the activity of A domain, the phosphate release assay was performed using 150 µM S3P, 0.2 mM PEP, 0.2 mM MESG 1U of PNPP and 10 nM AROM. The reaction was studied at 30 °C as above. For studying the GLP based inhibition of A domain, AroA assay was performed as above in presence of 0.2 mM GLP.

All the above reactions were monitored in the Microplate Reader Synergy™ H4 Hybrid or the Synergy™ Mx Monochromator-Based Multi-Mode Microplate Readers, and were carried out in triplicates. Further, the above experiments also included a no enzyme blank for each concentration, and a blank without any substrate.

2.7.3 Crosslinking Mass spectrometry

1 mg/ml AROM complex was crosslinked with 1 mM and 250 μ M of disuccinimidyl suberate. The digestion was carried out with endoprotease Lys-C followed by trypsin. The samples were measured on an Orbitrap Elite mass spectrometer, and the data was evaluated with the software xQuest [199].

2.7.4 Small Angle X-Ray Scattering (SAXS)

SAXS experiments were conducted at the beamline B21, Diamond Light Source (Didcot, UK), with X-ray wavelength of 1 \AA and a PILATUS 2M detector at a distance of 3.9 m. The samples contained 50 μ l of either Ct_AROM or Th_AROM at concentrations of 37 mg/mL, 15 mg/mL and 7mg/mL both as apo proteins and in the presence of various ligands (list of ligands given in Table 2.7).

Samples were delivered at 20 $^{\circ}$ C by an in-line Agilent HPLC with a Shodex Kw-403 column. The running buffer consisted of 20 mM HEPES pH 8.0, 150 mM NaCl, 2 mM TCEP and 1% sucrose. The continuously eluting samples were exposed for 300 s in 10 s acquisition blocks, and the data was pre-processed using in-house software.

Frames recorded immediately before the elution of the sample were used for the buffer subtraction. Buffer subtraction and further analysis were performed with Primus [195] and ScÅtter version 2.2b [196]. The crystallographic model of Ct_AROM was fit to the resulting SAXS profiles using the program FoXS [197].

Table 2.7 Ligands employed in the SAXS studies

Ligands (10 mM)
SKM
DHS
DHQ
ATP
NADH
SKM+ATP
SKM+NAD
DHQS+DHS
SKM+NADH
DHS+DHQ+SKM
DHS+NADH+ATP
SKM+NADATP
DHS+NADH+ATP
SKM+ATP+NADH+DHQS
SKM+ATP+NADH+DHQS+DHS

2.7.5 Negative stain Transmission Electron Microscopy screening

The Negative stain Transmission Electron Microscopy (NS-TEM) was carried out using glow-discharged carbon-coated grids. The grids were incubated with 0.1 mg/ml protein solution, stained with 1% uranyl acetate and examined with a FEI Tecnai G2 Spirit BioTwin transmission EM at 120 kV.

2.7.6 Crystallization and crystal handling

Various crystallization conditions for both Ct_ AROM and Th_ AROM were tested. The screen condition of Morpheus gave diffracting crystals. The resulting crystals were first optimized in grid screens and then in hanging drops. The final optimized conditions, together with the original Morpheus condition for Ct_ AROM samples are given in Table 2.6. The AROM crystals usually grew within two weeks. (Table 2.8)

Table 2.8 Crystallization conditions of Ct_ AROM

Sr.no.	Protein name	Protein Buffer	Conc mg/ml	Screen condition (SC)	Ligands 10mM	Cryo
1	Ct_ AROM (sitting drop)	20 mM Hepes pH 8.0, 150 mM NaCl, 2 mM TCEP	5	Morpheus H2 Buffer 1 pH: 6.5, 0.10 M aa, 0.03 mM P2: 60 %v/v	SKM, DHQ, NADH	SC
2.	Ct_ AROM (hanging drop)	20 mM Hepes pH 8.0, 150 mM NaCl, 2 mM TCEP	5	Buffer 1 pH: 6.5, 0.10 M aa, 0.03 mM P2: 60 %v/v	SKM, DHQ, NADH	SC
3.	Ct_ semet_ AROM (hanging drop)	20 mM Hepes pH 8.0, 150 mM NaCl, 2 mM TCEP	5	Buffer1 pH: 6.5, 0.10 M aa, 0.03 mM P2: 60 %v/v	SKM, DHQ, NADH	SC

2.7.7 Data collection and crystal structure determination AROM complex

Single wavelength Anomalous dispersion (SAD) data were collected at 100K and a wavelength of 0.978 Å on Pilatus 6M detector at beamline P14/PETRAII (Hamburg Germany). The data were indexed, integrated and scaled to 3.0 Å resolution using XDS. For heavy atom location, we employed SHELXD [198].

Most of the 78 expected heavy atoms corresponding to two chains of AROM in the ASU were located, with a prominent drop in peak height after about 75 sites. After phasing, density modification and chain tracing with SHELXE, large extents of the AroB and AroA domains and several secondary structure elements of the other domains became apparent. The superimposition of individual domains from the PDB aided in completing the structure, which was accomplished by cyclic chain tracing with Buccaneer [199], manual modeling with Coot [183], and refinement with REFMAC5 [182].

Table 2.9 Data collection and refinement statistics of Ct_AROM

Data set	Ct_AROM
Wavelength (Å)	0.978
Space group	P 21 21 2
Unit cell (Å)	a = 153.943, b = 377.616, c = 70.882, $\alpha = \beta = \gamma = 90.00^\circ$
A. Diffraction statistics	
Resolution	3.00
No. of observations	1122232 159766
Completeness	99.7 (99.3)
Redundancy	7.04
I/sigma(I)	13.67 (1.28)
R-meas (%)	11.2 (144.8)
CC _{1/2}	99.9 (54.1)
B. Refinement statistics	
Monomers/ASU	2
Resolution range (Å)	50 – 3.0
No. of unique reflections	79855
R _{cryst} (%)	21.5
R _{free} (%)	24.0
Rms bond length deviation	0.02
Rms bond angle deviation	0.753
Ramachandran statistics (%)	95.70/3.32/0.98

2.7.8. Conformational space Ct_AROM

To find structural representatives of the individual domains of AROM, the following five Pfam (version 31.0) families were used as starting point: AroB (PF01761), AroA (PF00275), AroK (PF01202), AroD (PF01487), and AroE (PF08501). The crystal structures were manually sorted based on the resolution and the structure with the highest resolution was picked as the reference structure. The reference structures are listed in Table 2.9

The structures were then clustered based on their differences in the opening angles between the N- and C-terminal (Nt and Ct) sub-domains (for AroB, A, K, and E). Further, each pair of structures from the same family was superposed first according to the alignment between their Nt sub-domains and then according to the alignment between their Ct domains. The rotation matrix during the 2nd transformation is used to calculate an “open angle difference” between the Nt and Ct sub-

domains of the two structures. TMalign is used to perform structure alignments. Each pair of structures in the same family was compared, and an “open angle difference” matrix was calculated accordingly.

A hierarchical agglomerative clustering with complete linkage was used to cluster the structures for each family. A cutoff of 0.15 rad (for AroB, A, and K) or 0.25 rad (for AroE), or 2.0 Å was used to acquire the clusters. The structures closest to the cluster centers were chosen as cluster representatives (medoids). R was used to perform the clustering. The representative structures for each of the four families are listed in the Table 2.10.

Table 2.9 Reference structures for each of the four flexible domains of AROM: The table shows the reference structures selected for each of the four domains based on the best resolution. The structures are manually divided into Nt and Ct sub-domains.

Domain	PDB ID	Resolution (Å)	Chain ID	start-end	Reference
AroB	1sg6	1.70	A	4-183, 184-355	[200]
AroA	3nvs	1.02	A	20-242, 6-19+243-421	[201]
AroK	2iyv	1.35	A	28-95, 11-27+96-167	[133]
AroE	3jyo	1.00	A	2-108, 109-283	[202]

Table 2.10. The representative structures for each of the four families.

<p>AroB (cutoff = 0.15 rad, num = 4) 1xal_B_1_317.pdb 1nva_B_1_356.pdb 5eks_B_1_323.pdb 3qbd_B_1_320.pdb</p>
<p>AroA (cutoff = 0.15 rad, num = 7) 1eps_A_6_420.pdb 1g6t_A_6_420.pdb 1rf5_B_4_421.pdb 2bjb_A_7_419.pdb 2gg4_A_12_440.pdb 3roi_A_6_421.pdb 4gfp_A_6_421.pdb</p>
<p>AroK (cutoff = 0.15 rad, num = 5) 3vaa_C_10_172.pdb 4y0a_A_27_184.pdb 1zyu_A_11_167.pdb 2g1j_A_11_167.pdb 3mrs_A_10_162.pdb</p>
<p>AroE (cutoff = 0.25, num = 7) 1vi2_B.pdb 3toz_H.pdb 1wxd_B.pdb 2ev9_A.pdb 2hk9_C.pdb 3pwz_A.pdb 3phi_A.pdb</p>

2.8. Methods- other proteins

2.8.1 Cloning, Protein Expression and Purification

Genomic DNA of *E.coli* was used as the template to clone the genes AroB, AroD, AroE, AroL, AroK, AroA in pET28b with the restriction sites NcoI and XhoI (C terminal hexa-His tag). *E. coli* enzymes were expressed at 37 °C. The cells were lysed using French press. The proteins were dialyzed in the buffer containing 10/20 mM Tris (pH 8.0), 150 mM NaCl, 2 β -ME. Gel filtration was carried out using S 200, 26/60. The *E.coli* enzymes contained, Leu and Glu prior to the His tag.

2.8.2. Kinetic Analysis of the *E. coli* enzymes

In order to make a comparison between the AROM complex and the monofunctional *E. coli* enzymes, the AroBDE kinetics was studied with the AroB, AroD, AroE, AroL, AroA enzymes from *E. coli*.

The 25 μ L reaction mixture contained 50 mM HEPES (pH 8.0), 100 mM KCl, 10 mM MgCl₂, 0.2 mM NAD, 0.2 mM CoCl₂, 0.8 mM NADPH, 0.8 mM ATP, 0.8 mM PEP. Each of the *E. coli* enzymes (AroB, AroD, AroE, AroL, AroA) was supplemented at a concentration of 10 nM. The DAHP range was the same as in section 2.7.2. The reaction was monitored at 30 °C in the same way as described for the AroBDE kinetics of AROM. (Section 2.6.6).

3. Results and Discussion

3.1 Cluster analysis of the shikimate kinase fusions

Shikimate pathway enzymes show substantial variance in the molecular and structural organization within different taxonomic groups [49]. In most prokaryotes, the seven enzymes are encoded as separate well characterized single proteins [39]. Plants encode all monofunctional enzymes except one bifunctional fusion protein AroDE [33, 34]. Fungi encode the pentafunctional AROM complex (AroBAKDE) [49, 79, 161, 167, 203-207], which was also reported in apicomplexan parasites [36, 165]. Acanthamoeba, on the other hand, encodes a tetrafunctional AroN lacking the E domain (AroBAKD) [37]. There could be many yet unidentified fusion proteins in the pathway. Therefore, a detailed study of the evolutionary origins of shikimate pathway is needed.

Richards *et al.*, performed such a study in 2006. They built phylogenetic trees of AroG, AroB, AroE, AroK, AroA and AroC [35]. They detected previously unknown AroKB in various classes of both gram-positive and gram-negative bacteria. They further identified AroBD from chromalveolates and AroDEA from chlamydiae. Although, each of these calculations was based on less than 100 sequences, the authors already found a variety of previously unreported multifunctional enzymes in the shikimate pathway.

Today, with a vast repertoire of new genomic data available, a new homology search might uncover even more yet undiscovered multifunctional enzymes. Therefore, at the beginning of this project, a homology search for AroK analogs in the nr-70 database was conducted. Approximately 5000 sequences were retrieved, 3655 AroKs were manually selected, and were clustered by all-against-all pairwise similarities as measured by the BLAST p-values to produce a 2-dimensional map in CLANS [176] (Figure 3.1).

Besides the evolutionary relationship of AroKs, this map also shows the different multifunctional enzymes, comprising the AroK domain. Various previously unreported multifunctional enzymes in the shikimate pathway can be seen in this map. The map shows the extent of the distribution of AroKBs (fuchsia stars) among various taxonomic groups. The multiple AroKBs scattered throughout the map indicate that AroKB fusion has evolved multiple times. In most of these

organisms, AroK and AroB genes are present adjacently in one operon, therefore, only the deletion of the stop codon is required for this fusion.

Interestingly, different fusions between the AroE and AroK domains (blue stars) were detected. AroKE fusion occurs in Archaea and AroEK fusion is present in bacteria. The bacterial AroEK forms two distant clusters, which suggests that, the AroEK fusion has evolved at least twice. The gene arrangement is different in AroEK and AroKE but it is unknown whether there are any differences between the domain arrangements of these proteins AroKE and AroEK. Further, a novel fusion between AroD and AroK (lemon stars) is also detected in firmicutes.

The AROM complex (purple stars), other than in fungi, is present in choanoflagellates (single-celled aquatic protists), oomycetes (water molds) and chromerida (photosynthetic alveolates, closely related to apicomplexan parasites) [208-210]. Further, the tetrafunctional AroN (lavender-blue star) in *A. castellanii* is also visible in the map.

The other fungal cluster, which does not represent the AROM complex, contains the quinate repressor protein (QutR, Uniprot ID: P11637) (bazaar circles). QutR contains homologs to the three C-terminal domains (KDE) of the pentafunctional AROM complex and functions as a molecular sensor that detects the presence of quinate/shikimate pathway intermediates to control transcription regulation [211, 212].

With the discoveries in this single AroK-based cluster map, it is apparent that significantly deeper and broader evolutionary insights might be obtained in future studies integrating cluster maps of the other enzymes of the pathway.

The presence of these multifunctional enzymes suggests the possibility of existence of a higher order assembly. Higher order assemblies are known to concentrate reactants to avoid unfavorable reactions, and remove inhibitory products thus channeling metabolites from one enzyme to the next [18, 19]. With such prominence of these fusions, it seems likely that the individual enzymes of the pathway form transient complexes for efficient substrate funneling and catalysis. It could further be that these transient higher order assemblies might resemble the AROM complex.

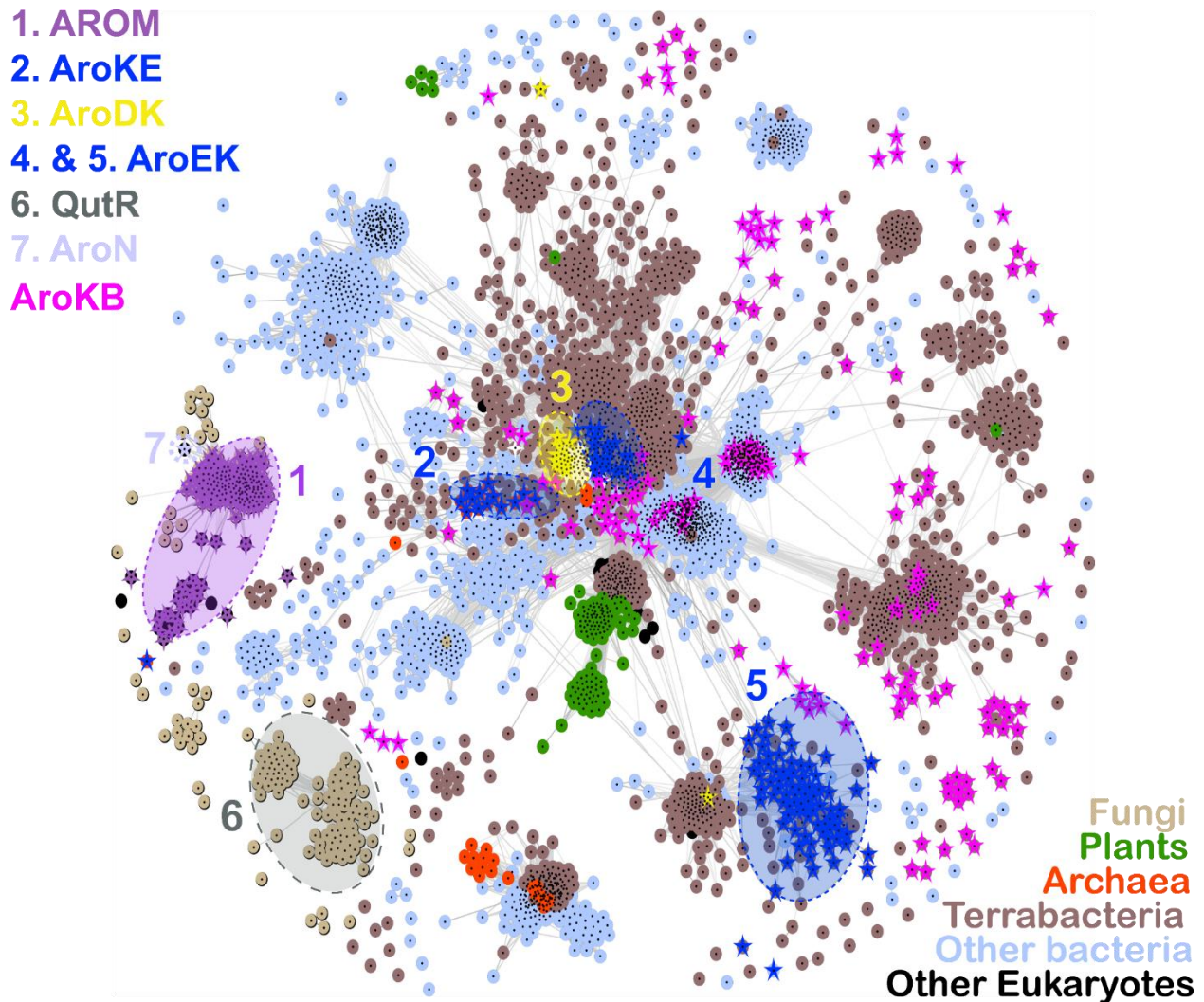


Figure 3.1 Cluster analysis of the shikimate kinase genes: The figure shows the cluster analysis of shikimate kinase genes. The sequences are represented by dots and the line coloring reflects BLAST p-values. The darker a line, the lower the p-value. The organisms, which are shown as circles include fungi, plants, archaea, terrabacteria, other bacteria and other eukaryotes and are colored with sorrel brown, green, red, bazaar, pale blue, and black respectively. The fusions, which are shown as stars include AroKB, AroKE, AroEK, AroDK, AROM and AroN are shown as stars and are colored with fuchsia, blue, lemon, purple, and lavender-blue respectively.

AroKB, AroKE and AroEK together with AroDE (PDBID: 2O7S) constitute four domains of the pentafunctional AROM complex (Figure 3.2). If one could determine the domain arrangements of these three bifunctional enzymes, then together with the AroDE, a tetrafunctional hypothetical model of the AROM complex can be designed *in silico*. Further, a kinetic analysis of these bifunctional enzymes can also give insights into allosteric regulation between these domains. Not only can such a study help us understand the shikimate pathway better, it can also strengthen our

knowledge of kinetic mechanisms *in vivo*. Therefore, a few candidates were selected from AroKB, AroKE and AroEK for further studies.

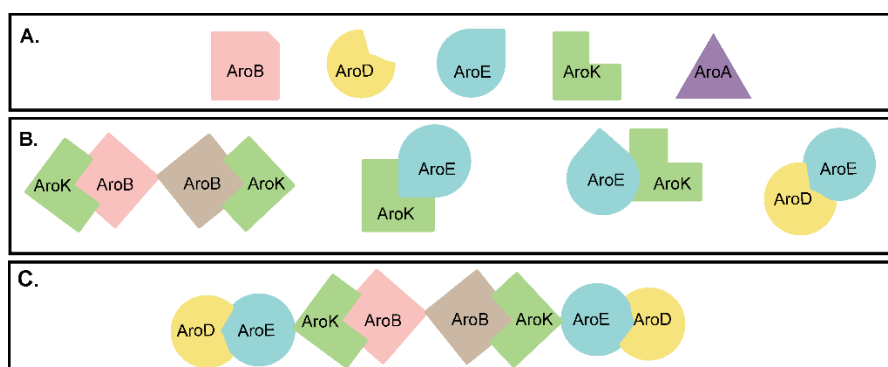


Figure 3.2. Real bifunctional enzymes and hypothetical model of AROM: The figure compares the monofunctional enzymes of the pathway with the bifunctional enzymes and shows the hypothetical tetrafunctional hypothetical model. A). Five monofunctional enzymes constituting the pentafunctional AROM complex. B). Three bifunctional enzymes selected for this study, AroEK, AroKB, and AroKE together with AroDE from literature C). AroKB, AroEK/AroKE together with At_AroDE constitute four domains of the pentafunctional AROM complex, with only A domain missing, therefore, a hypothetical tetrafunctional model of AROM complex can be designed.

3.2. AroKB

The AroKB fusion (Figure 3.3), which involves the second and the fifth enzyme of the pathway, occurs very frequently in different taxonomic groups as seen in the CLANS map (Figure 3.1). It seems not catalytically efficient if the substrate diffuses out of AroKB for reaction three and four, only to catalyze reaction five again by AroKB. Therefore, a channeling effect might not be feasible for Tm_AroKB but a regulation between the two domains might regulate the pathway. In addition, this fusion could suggest the presence of potential higher order assembly in the cell containing AroKB. In this section the purification, biochemical and biophysical properties, kinetic analysis and the crystal structural of bifunctional AroKB is discussed. This is further compared to the monofunctional Tm_AroK_{KB} and Tm_AroKB_SB.

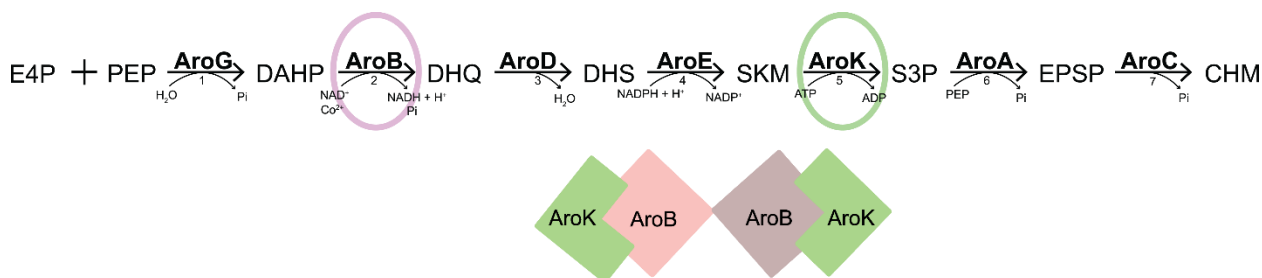


Figure 3.3 Bifunctional AroKB: The figure shows a hypothetical model of bifunctional AroKB containing the fusion of the second and fifth steps of the shikimate pathway.

3.2.1 Biochemical and biophysical characterization

AroKBs from five different organisms were recombinantly expressed in *E. coli*. AroKB from *Collimonas fungivorans* did not express and AroKBs from *Acidiphilium multivorum*, *Granulibacter bethesdensis* CGDNIH1, and *Oxalobacter formigenes* were insoluble. Initial optimization did not result in a soluble protein (data not shown). Only the AroKB from *Thermotoga maritima* (Tm_AroKB) yielded soluble protein. *T. maritima* is a strictly anaerobic, rod-shaped, fermentative, and hyperthermophilic [HT] eubacterium. It grows between 55 and 90 °C, with an optimum temperature of around 80 °C [213]. The purified protein is well folded with a thermal melting temperature (T_m) of 85 °C as seen by CD. It migrates as a single species in size exclusion chromatography with an average molecular mass of ~116 kDa as determined by static light scattering (Figure 3.4).

A multifunctional enzyme is best studied if a truncated version of the same enzyme is available for comparison. Therefore, Tm_AroK_{KB}, which is a 159 amino acid truncated protein containing only the monofunctional K domain was also cloned, expressed and purified to study the effect of absence of the B domain. The purified protein is a 22 kDa monomer. It is a well-folded protein with a thermal melting temperature (T_m) of 70 °C as seen by CD. (Figure 3.5). It is therefore slightly less stable than the bifunctional Tm_AroKB. Therefore, it appears that the presence of the second domain has a stabilizing effect on the enzyme. Interestingly, a truncated protein containing only the B domain could not be obtained in folded state.

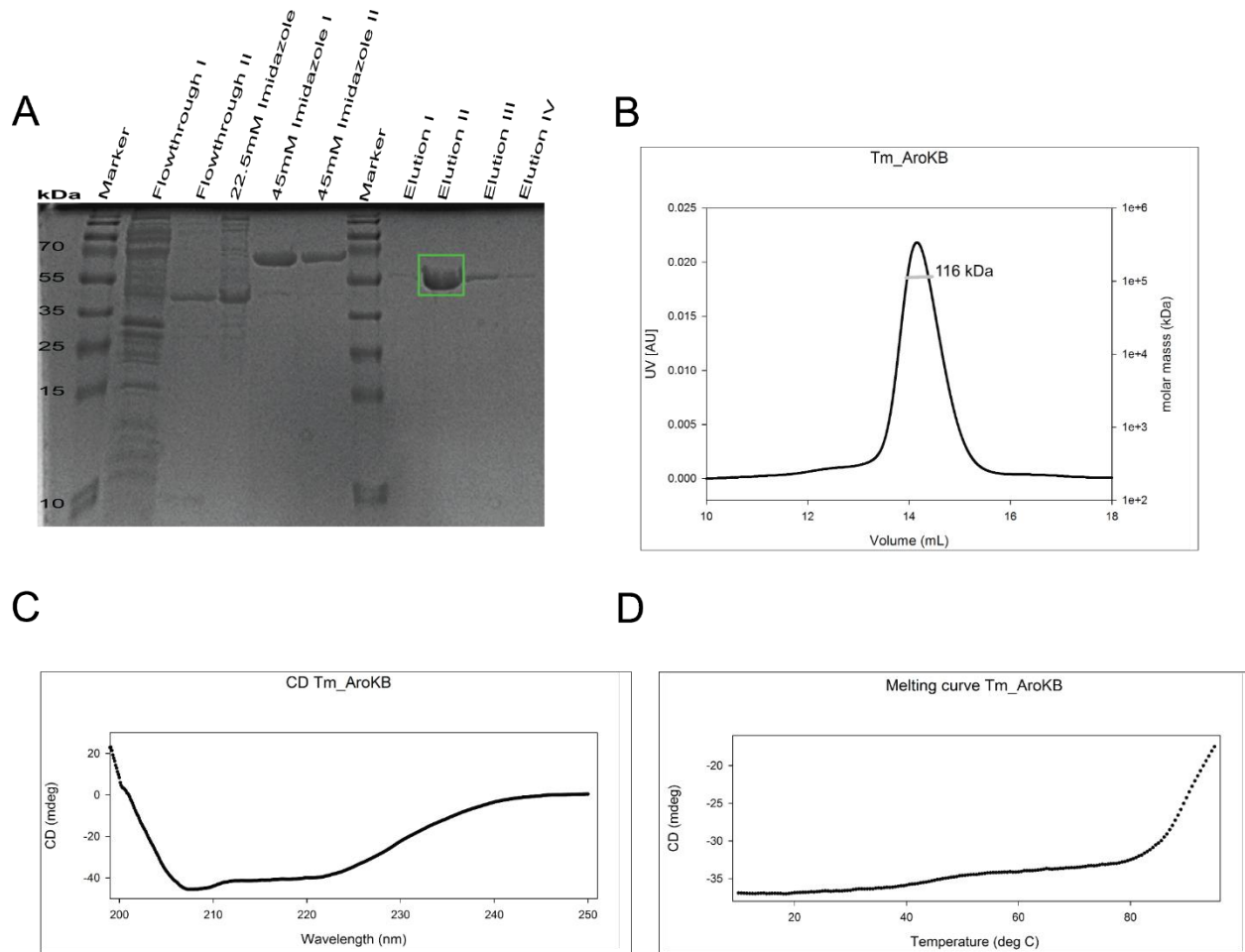


Figure 3.4 Biochemical and biophysical properties of Tm_AroKB: **A). SDS gel of Tm_AroKB:** The gel shows the molecular weight of the monomer of the AroKB is 56 kDa. 10 μ l of samples were loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B). Light scattering of Tm_AroKB:** 100 μ l Tm_AroKB was loaded at a concentration of 1 mg/ml (black profile), onto a S200 size exclusion column and the mass of the eluted particles (gray profile) in the peak area was analyzed with static light scattering. The horizontal axis of the curve shows volume in mL and the left axis shows the UV absorption. The molar mass on the right is shown in the logarithmic scale. **C). CD profile of Tm_AroKB:** 400 μ l of Tm_AroKB at a concentration of \sim 0.3 mg/mL was measured in the CD spectrometer. The figure shows the far UV CD spectrum, plotted in millidegrees (mdeg) versus wavelength in nanometers (nm) of polarized light. **D) Thermal stability analysis of Tm_AroKB:** 400 μ l of Tm_AroKB at a concentration of \sim 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 85 $^{\circ}$ C.

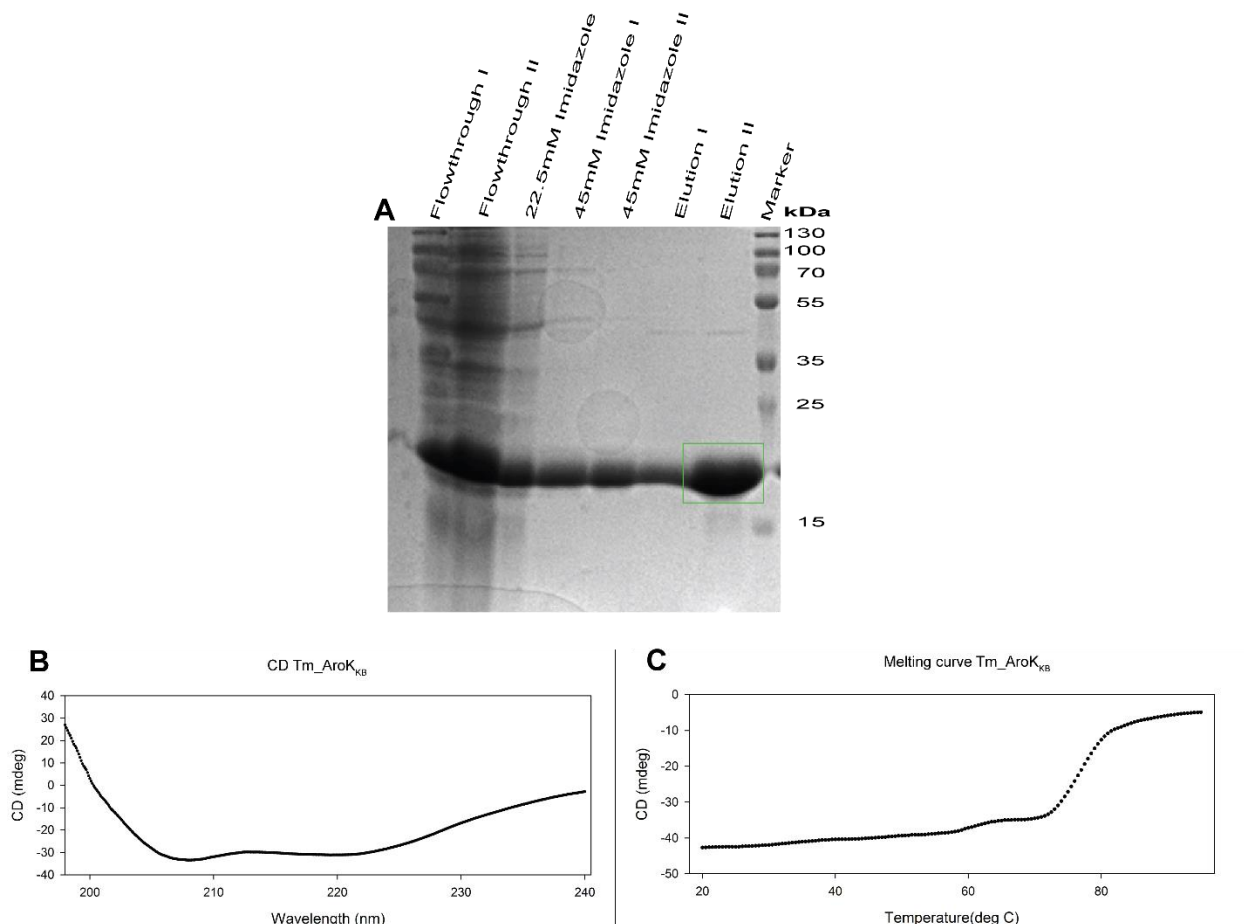


Figure 3.4 Biochemical and biophysical properties of Tm_AroK_{KB} **A).** SDS gel of Tm_AroK_{KB}: The gel shows the molecular weight of the monomer of the AroKB is 22 kDa. 10 μ L of samples were loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B).** CD profile of Tm_AroK_{KB}: 400 μ L of Tm_AroK_{KB} at a concentration of -0.3 mg/mL was measured in the CD spectrometer. The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarized light. **C) Thermal stability analysis of Tm_AroK_{KB}**: 400 μ L of Tm_AroK_{KB} at a concentration of -0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 70 $^{\circ}$ C.

3.2.2 Kinetic analysis of Tm_AroKB

The activities of both the K and B domains of Tm_AroKB were assayed separately in presence and absence of the substrates of the other domain. Due to its thermophilic nature, the true potential of the enzyme may possibly only be realized around 80 $^{\circ}$ C. However, this temperature cannot be employed for the reaction due to the stability of coupling enzymes.

For the assay of K domain, the release of ADP is coupled to the pyruvate kinase (PK) (EC 2.7.1.40) and lactate dehydrogenase (LDH) (EC 1.1.1.27) (PK/LDH) enzymes (Figure 3.6). The SKM/ATP dependent oxidation of NADH is in turn monitored at 340 nm ($\epsilon = 6180 \text{ M}^{-1} \text{ cm}^{-1}$) [185]. The

coupling enzyme cocktail of PK/LDH is isolated from rabbit muscle, and therefore denatures above 40 °C (data not shown). K domain kinetics was studied for both substrates SKM and ATP. To examine the effect of an increase in temperature, K domain kinetics of Tm_AroKB was conducted both at 30 and 40 °C. The rate of the reaction appears to increase slightly with the increase in temperature (data not shown). Therefore, it could be a possibility that the enzyme is actually more active than seen here at 30 °C.

During the formation of DHQ by the B domain, an inorganic phosphate is released. The AroB kinetics was followed by monitoring the continuous phosphate release employing the Enzcheck® phosphate release kit [188] and measuring the amount of phosphorylated MESG at 360 nm (Figure 3.7). The released phosphate is transferred to 2-amino-6-mercapto-7-methylpurine riboside (MESG), which is in turn converted to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine by purine nucleoside phosphorylase (PNP). The accompanying change in absorption at 360 nm allows quantification of inorganic phosphate (Pi) released in the reaction. To estimate the extinction coefficient of phosphate ($\epsilon = 14753 \text{ M}^{-1} \text{ cm}^{-1}$) a standard curve was prepared; using a phosphate concentration range of 4.3 to 150 μM in duplicates (data not shown). MESG is not stable beyond room temperature. Therefore, the AroB assay was only carried out at 30 °C. The kinetics for B domain was conducted both for DAHP and NAD.

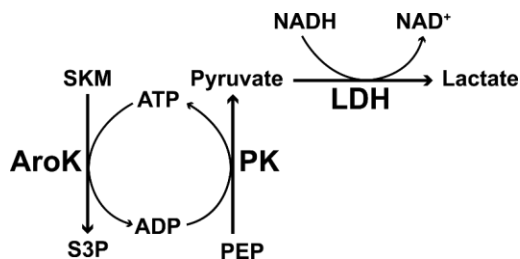


Figure 3.6 Assay of K domain: The standard method for kinase assay. The method couples the release of ADP to PK and LDH and measures the ADP dependent consumption of NADH at 340 nm.

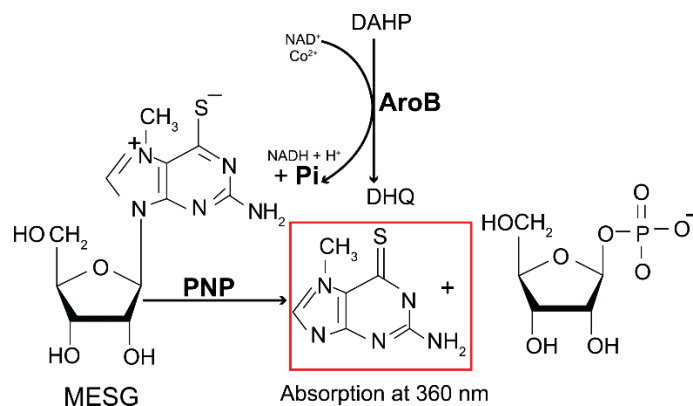


Figure 3.7 Assay of B domain: In this method 2-amino-6-mercapto-7-methylpurine riboside (MESG) is converted to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine by purine nucleoside phosphorylase (PNP). The accompanying change in absorption at 360 nm allows quantification of inorganic phosphate (Pi) consumed in the reaction.

Separate experiments were conducted for both the K and B domains of the enzyme in the presence of the substrates of the other domain to test for potential allosteric effects; i.e. to measure the effect of the presence of the substrates of B domain on K kinetics and vice versa. The experiments for both the K and B domains were conducted also for the salt bridge mutant Tm_AroKB_SB, which does not contain the bound NAD present in Tm_AroKB (discussed in section 3.2.4).

Table 3.1 lists the estimated values of the kinetic parameters for each of the different domains of Tm_AroKB in comparison to the values reported for monofunctional enzymes in the literature. The MM curves for kinetic analysis of the K domain of KB are presented in Figure 3.8. The analysis of the K domain of Tm_AroKB was conducted for both substrates, SKM and ATP. The k_{cat}/K_m falls in the same range, as reported for *M. tuberculosis* (Table 3.1) [131, 187] which shows that at 30°C, Tm_AroKB is at least as active as enzymes reported in the literature. The analysis of Tm_AroKB was also conducted at 40 °C; it showed a slight increase in the activity (data not shown). Therefore, it could be highly likely that the enzyme is much more active than seen in these experiments but its true potential cannot be realized at this temperature.

In comparison, the activity of monofunctional Tm_AroK_{KB} does not change much in the absence of the B domain. The kinetic parameters are almost the same as seen for the bifunctional Tm_AroKB. Therefore, a difference in activity of the monofunctional enzyme compared to that of bifunctional enzyme was not observed. Thus, a regulatory effect of the presence of other domain was not visible.

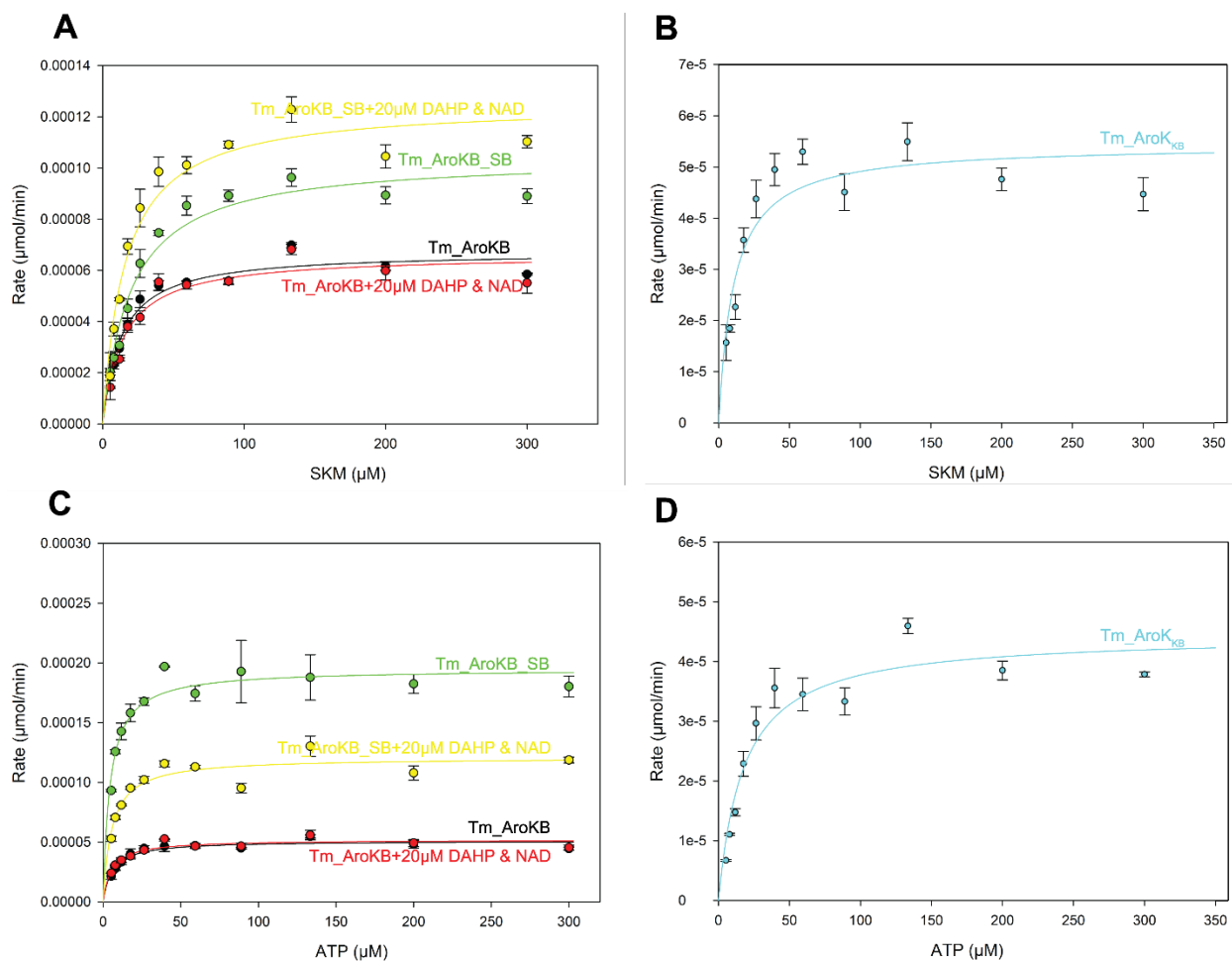


Figure 3.8 K domain kinetics of Tm_AroKB: The Figure shows the kinetic analysis of the K domains of Tm_AroKB, Tm_AroKB_SB and Tm_AroK_{KB}. The enzyme concentration for Tm_AroKB and Tm_AroKB_SB is around 10 nM and for Tm_AroK_{KB} is around 4.5 nM. The enzyme names are indicated in the graphs together with their respective curves. **A). Tm_AroKB K domain kinetics for SKM:** This panel compares the SKM kinetics of the K domain between Tm_AroKB and Tm_AroKB_SB in presence and absence of substrates of the B domain. **B). Tm_AroK_{KB} K domain kinetics for SKM:** This panel shows the SKM kinetics of Tm_AroK_{KB}. **C). Tm_AroKB K domain kinetics for ATP:** This panel compares the ATP kinetics of the K domain between Tm_AroKB and Tm_AroKB_SB in presence and absence of substrates of the B domain. **D). Tm_AroK_{KB} kinetics for ATP:** This panel shows the ATP kinetics of Tm_AroK_{KB}. Due to differences in protein concentration, the curves for Tm_AroKB and Tm_AroKB_SB are not directly comparable.

The kinetic analysis of the K domain was also carried out in the presence of substrates of the B domain to assess the effect of the presence of other substrates. The activity of Tm_AroKB was assessed in the presence of 20 μM DAHP and NAD.

Tm_AroKB contains a bound NAD, which could potentially increase the catalytic turnover of the enzyme. To study this effect, a mutant Tm_AroKB_SB, which does not contain this bound NAD was designed (section 3.2.4). The kinetic activities for both domains of Tm_AroKB were

compared to Tm_AroKB_SB. At first glance, Tm_AroKB_SB seems to be more active than that of Tm_AroKB for both substrates SKM and ATP. However, NAD is present in the binding pocket of Tm_AroKB, which influences the absorption at 280 nm. Due to this effect, the concentration of Tm_AroKB is potentially overestimated in comparison to Tm_AroKB_SB. As the increase in activity is not very significant, it can be attributed to this overestimation.

While the catalytic activity of Tm_AroKB is unaffected in the presence of the substrates of the B domain (NAD and DAHP), the activity of Tm_AroKB_SB is decreased. This could be explained by a competition between NAD and ATP for the ATP binding site [214]. Therefore, it could be that intrinsic binding of NAD in Tm_AroKB is to avoid this competition, rather than to increase the catalytic efficiency of the enzyme. The activity of the K domain was determined only in the presence of the two substrates of B domain was not carried out separately in this study.

The MM curves for kinetic analysis of the B domain of Tm_AroKB are presented in Figure 3.9. The kinetic analysis of the B domain was conducted for both DAHP and NAD. To our surprise, Tm_AroKB, which contains an already bound NAD, showed only negligible activity with NAD as a substrate (data not shown). On the other hand, Tm_AroKB_SB, which does not contain bound NAD, showed kinetic activity towards NAD. As seen in the Table 3.1, the kinetic parameters fall in the range of the kinetic values reported in the literature for *A. chinensis* and *A. nidulans* [215, 216].

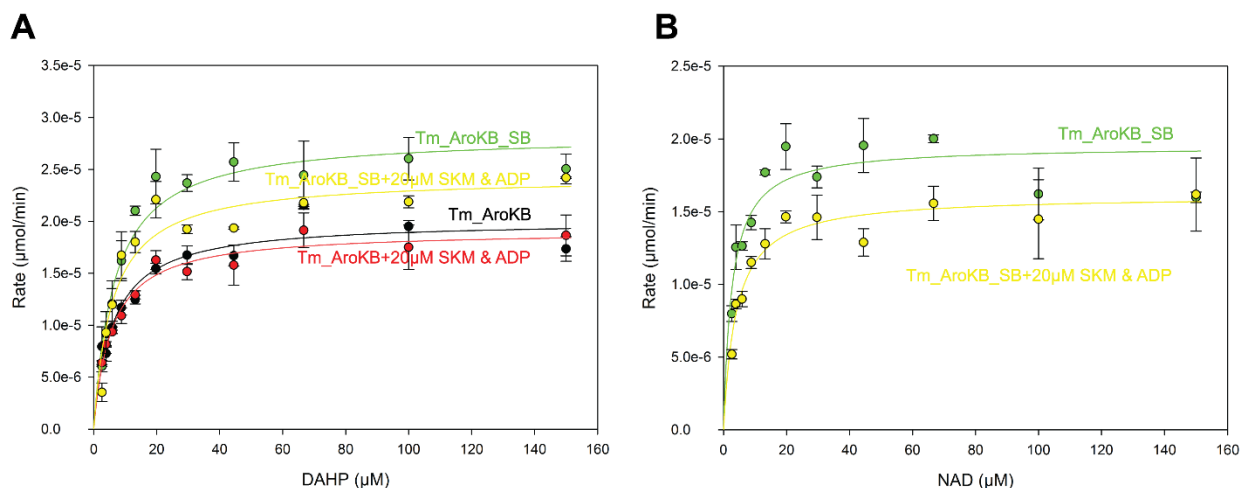


Figure 3.9 B domain kinetics of Tm_AroKB: The figure shows the kinetic analysis of the B domain of Tm_AroKB, and Tm_AroKB_SB. The enzyme concentration for Tm_AroKB and Tm_AroKB_SB is around 10 nM. The enzyme names are indicated in the graphs together with their respective curves. **A). B domain kinetics of Tm_AroKB for DAHP:** This panel compares the DAHP kinetics of the B domain between

Tm_AroKB and Tm_AroKB_SB in the presence and absence of substrates of the B domain. **B). B domain kinetics for NAD:** This panel shows the NAD kinetics of the B domain of Tm_AroKB_SB in the presence and absence of substrates of the K domain. Due to differences in protein concentration, the curves for Tm_AroKB and Tm_AroKB_SB are not directly comparable.

In earlier studies, the kinetic analysis of B domain was coupled to the D assay. This was not done here, because the requirement of this study is to analyze the kinetics of the B domain in the presence of substrates of the K domain. The reason for not choosing the D coupling is the phosphate buffer employed usually for the D kinetics. Phosphate buffer can interact with the ATP binding site in K domain because it can bind instead of ATP at the p-loop [133]. Therefore, the analysis of the B activity in presence of SKM and ATP cannot be carried out if the assay was carried out in phosphate buffer.

The activity of DAHP was studied for both the B domains of Tm_AroKB and Tm_AroKB_SB. The activity of Tm_AroKB_SB was found to be almost 20% higher than that of Tm_AroKB. However, as stated previously, the NAD bound to Tm_AroKB contributes to the absorbance of Tm_AroKB, leading to an overestimation of the concentration of Tm_AroKB. Therefore, conclusions are to be drawn with care. However, the addition of the substrates for the K domain clearly reduces the activity for DAHP of the B domain. This could be attributed to the ATP binding competing with NAD-binding [214].

The NAD activity was also tested for both Tm_AroKB and Tm_AroKB_SB. Tm_AroKB did not show a significant change with increasing concentration of NAD, this could be attributed to the fact it already contains bound NAD and therefore, increasing concentrations of NAD do not make any difference (data not shown).

In the light of these results, a conclusion about the regulatory interactions between the two domains cannot be established.

Table 3.1 Kinetic parameters of Tm_AroKB: The table shows the kinetic parameters of Tm_AroKB obtained in this study and compares them to the values in literature. TS refers to this study.

Enzyme	Substrate (S) (μM)	Substrate (other domain) ($20 \mu\text{M}$)	$K_m \pm \% \text{ S.E.}$ (μM)	E_t (nM)	$k_{cat} \pm \% \text{ S.E.}$ (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \text{ M}^{-1}$)	R^2	Ref
Tm_AroKB	SKM	-	12.85 \pm 10.75%	10	1.122 \pm 2.6%	8.73e+4	0.916	TS
		DAHP, NAD	14.19 \pm 13.18%		1.101 \pm 3.2%	7.76e+4	0.887	
	ATP	-	5.98 \pm 15.47%		0.851 \pm 2.76%	1.42e+5	0.797	
		DAHP, NAD	5.43 \pm 14.88%		0.866 \pm 2.65	1.59e+5	0.777	
	DAHP	-	6.01 \pm 14.37%		0.334 \pm 3.4%	5.56e+4	0.839	
		SKM, ADP	5.71 \pm 16.62%		0.318 \pm 3.79%	5.57e+4	0.808	
	NAD	-	-		-	-	-	
Tm_AroKB_SB	SKM	-	19.91 \pm 10.9%	10	1.737 \pm 3.01 %	8.72e+4	0.933	TS
		NAD	15.65 \pm 12.75%		2.085 \pm 3.28 %	1.33e+5	0.904	
	ATP	-	4.46 \pm 16.7%		3.245 \pm 2.72 %	7.28e+5	0.72	
		DAHP, NAD	5.54 \pm 14.08%		2.01 \pm 2.5 %	3.63e+5	0.800	
	DAHP	-	6.57 \pm 16.05%		0.472 \pm 4.00 %	7.18e+4	0.845	
		SKM, ADP	5.86 \pm 16.62%		0.404 \pm 3.92%	6.89e+4	0.827	
	NAD	-	2.68 \pm 19.6%		0.325 \pm 3.62 %	1.21e+5	0.736	
		SKM, ADP	3.99 \pm 19.2%		0.268 \pm 4.07%	6.72e+4	0.691	
Tm_AroK _{KB}	SKM	-	10.67 \pm 18.79%	4.6	1.595 \pm 4.2%	1.49e+5	0.777	TS
	ATP	-	17.70 \pm 15.83%	4.6	1.953 \pm 4.3%	1.10e+5	0.868	
Mtb_AroK	SKM	-	650 \pm 4.3%, 410 \pm 4.87%	-	-	9.0e+4, 1.1e+5	-	[131, 187]
	ATP	-	112 \pm 0.03%, 83 \pm 4.8%	-	-	5.4e+5 5.3 e+5	-	
Ac_AroB	DAHP	-	3.2 \pm 6.5%	-	0.50 \pm 2%	1.56e+5	-	[215]
An_AroB	DAHP	-	21 \pm 9.5%	-	6.8	3.24e+5	-	[216]
	NAD	-	1.9 \pm 5.3%		-	-	-	

3.2.2 Crystal structure of Tm_AroKB in complex with SKM

Initial crystals of Tm_AroKB were obtained in complex with SKM. Initial crystals diffracted to approximately 3 Å resolution. With grid screens, the resolution improved to 2.9 Å. The crystal structure of 1DQS was used as the search model for the AroB domain, while 2IYR was used as the search model for the AroK domain for molecular replacement. The asymmetric unit contained only one Tm_AroKB molecule. The two constitutive domains of the monomer are found to be connected via a rigid interface, and the active sites are solvent accessible (Figure 3.10). For the catalytic activity, the B domain always exists as a dimer in solution [61, 91, 109, 200, 216-221]. In the crystals, this dimer is constructed via crystallographic symmetry (Figure 3.5). The two K domains are connected at the outward side of the B domain dimer. Such an arrangement of the bifunctional enzyme does not affect the solvent accessibility of the active sites of both K and B domain. The protein is bound to the ligands NAD, Zn²⁺, SKM and SO₄²⁻. As the protein was co-crystallized together with SKM, and ammonium sulfate was a component of the crystallization buffer, NAD and Zn²⁺ might have bound to the B domain within the cell.

The structure of the K domain contains bound SKM and sulfate. SKM binding is similar to previously reported structures. The crystallization buffer contained sulfate, which can bind instead of phosphate at the p-loop and may occupy a location that is nearly identical with the position of the β-phosphates of ATP or ADP relative to the p-loop.

The architecture of the B domain of Tm_AroKB is closely similar to the B domains from other organisms. Each monomer of the B domain is composed of two subdomains, the N-terminal α/β domain, containing a Rossmann-fold, and serves as NAD-binding site and the α-helical C-terminal domain that contains most of the residues involved in catalysis, including the Zn²⁺ binding site. Formation of each of the two active sites within the dimer requires the interaction of amino acid residues from domains N and C of one monomer [61]. Therefore, the side chain of Arg-262 extends into the other monomer for catalysis, which is marked in Figure 3.10.

3.2.3 Tm_AroKB_SB

No other AroB in the literature contains an intrinsically bound NAD [61, 91, 92, 200, 221]. NADH together with Zn²⁺ or Co²⁺ is required for catalytic activity of AroB [61, 218]. NADH can be regenerated from NAD during the reaction catalyzed by B domain [217]. Therefore, the presence of bound NAD/NADH could potentially enhance the catalytic turnover of AroKB. To confirm

whether the protein indeed contains bound NAD and not NADH, the absorption spectrum of the protein was recorded at 340 nm, but the protein did not show any absorbance at 340 nm (data not shown). Further, NAD binding could potentially stabilize the protein and contribute to thermostability. Therefore, unfolding and refolding of the protein to release the bound NAD was attempted. However, the protein could not be refolded successfully.

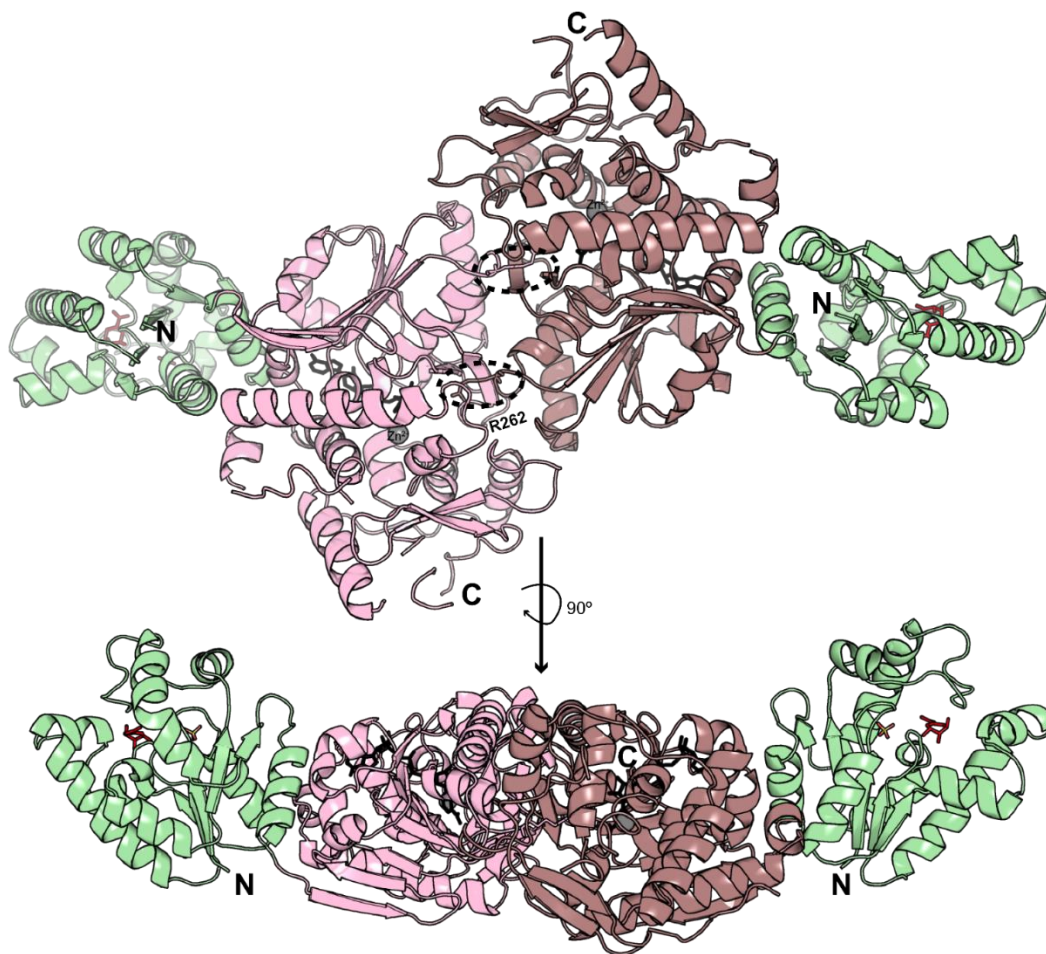


Figure 3.10 Crystal structure of Tm_AroKB in complex with SKM and NAD. The figure shows the dimer of bifunctional AroKB. The N and C termini are marked. The AroK domain is shown in green and the AroB domain is shown in light pink and dirty violet, to highlight the two dimeric domains. The Arg-262, extending in the other half of the dimer is shown as pink sticks (circled). SKM is shown as red sticks, while sulfate is shown as red and yellow sticks. The AroB domain contains NAD and Zn²⁺ bound, which are shown in black and gray respectively.

On comparing the crystal structure of Tm_AroKB with the *Aspergillus nidulans* AroB (An_AroB) in complex with Zn²⁺ and NAD [206], it was observed that the NAD binding pocket of Tm_AroKB was less solvent accessible. On a closer look, the NAD-binding pocket in Tm_AroKB was found to contain a salt bridge between Arg-195 and Asp-316 (Figure 3.11). Thermophilic proteins are

usually reported to have more salt bridges, which can help them achieve higher thermostability [222].

Herein, it was hypothesized that this salt bridge is aiding Tm_AroKB to adapt at higher temperatures, while also trapping the NAD the binding pocket, which could potentially stabilize the protein and aid catalysis. Thus, a mutant (Tm_AroKB_SB) which contains Ala-195 and Ala-316 instead of Arg-195 and Asp-316 was designed. If this salt bridge, and thus the less accessible binding pocket are contributing to NAD binding, then Tm_AroKB_SB should not contain intrinsically bound NAD. The purified protein was well folded, with a thermal melting temperature (Tm) of 85 °C (Figure 3.12). No change in the melting temperature suggests that the salt bridge between Arg-195 and Asp-316 does not contribute to the thermostability of this protein.

Unlike Tm_AroKB, Tm_AroKB_SB did not crystallize in complex with SKM. However, in complex with SKM and NADH or SKM, NADH and ADP, crystals were readily obtained under the same conditions as for Tm_AroKB. This shows that, NAD is required for the protein crystallization but does not affect the thermal stability of this protein. Further, this shows that the salt bridge indeed helps the protein in binding NAD tightly intracellularly.

The Tm_AroKB_SB crystals had the same space group and cell dimensions as the Tm_AroKB crystals, so the structure could be solved on the basis of the Tm_AroKB coordinates. Figure 3.13 shows the crystal structure of Tm_AroKB_SB containing Ala at positions 195 and 316. The more solvent accessible NAD-binding pocket can be seen in the crystal structure but NAD is bound at the same orientation. Thus, the salt bridge between Arg-195 and Asp-316 is not necessary for correct NAD binding, but helps entrapping NAD in the binding site.

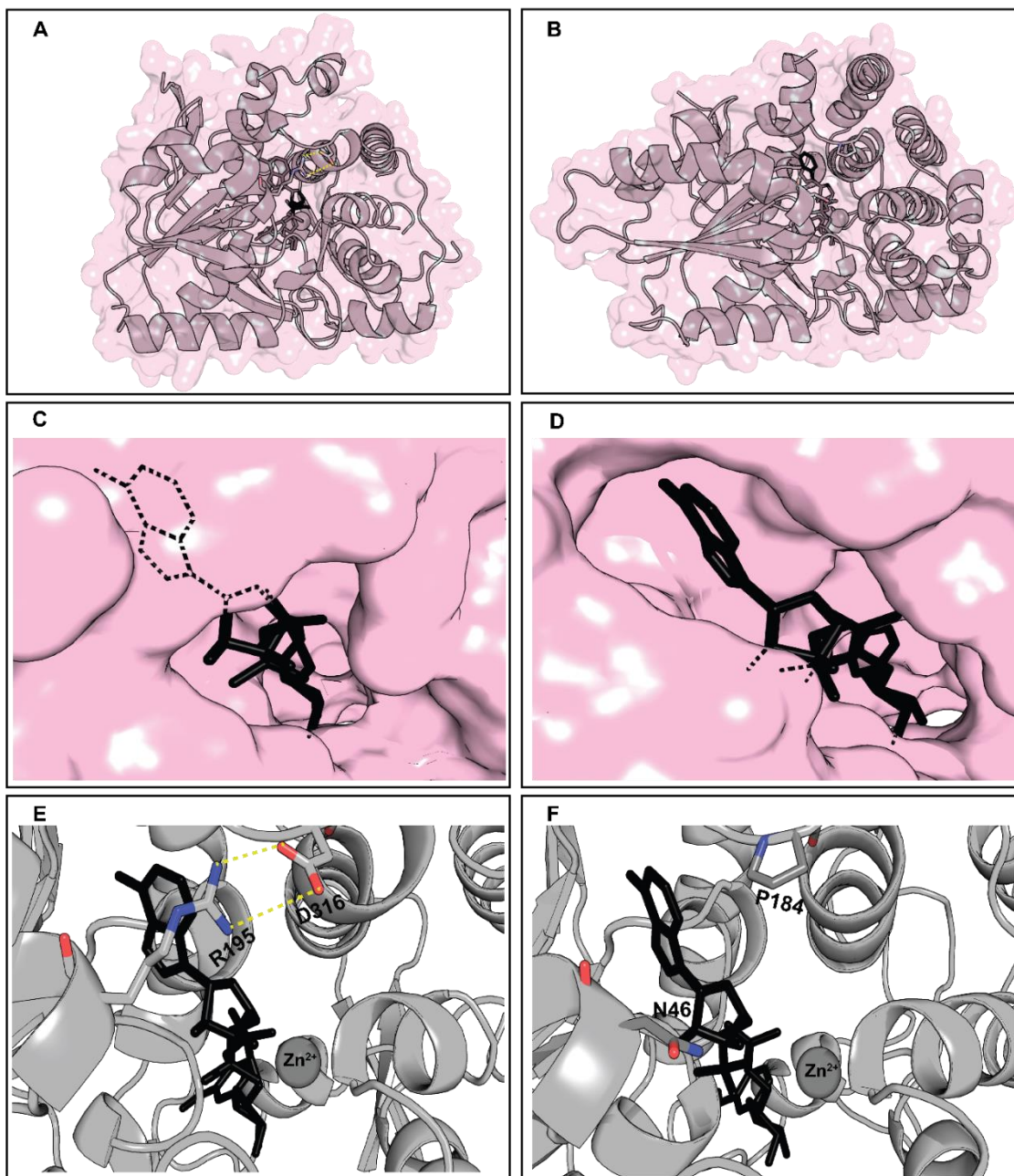


Figure 3.11 NAD-binding pocket: The Figure compares the NAD-binding pocket of Tm_AroKB with the NAD-binding pocket of *A. niger* AroB. Only one monomer of AroB is shown for both. The figure shows both proteins in combination with NAD and Zn²⁺. The surface is shown in pink color while the cartoon is shown in gray color. The salt bridge, which makes the binding pocket less accessible, is shown in yellow. **A) Transparent surface representation and cartoon representation of Tm_AroKB:** The panel shows the transparent surface representation of Tm_AroKB on top of its cartoon representation. The salt bridge, which makes the NAD-binding pocket less accessible, is visible in the cartoon representation. **B) Transparent surface representation and cartoon representation of An_AroB:** The panel shows the transparent surface representation of An_AroB on top of its cartoon representation. The solvent accessible pocket present in An_AroB is also seen. **C) Zoomed binding pocket surface of Tm_AroKB:** This panel shows a zoomed surface of NAD-binding pocket. NAD is shown within in the binding pocket as sticks and the part of NAD which is not solvent accessible is shown as dashed lines. **D) Zoomed binding pocket surface of An_AroB:** The panel shows the surface representation of NAD-binding pocket zoomed. NAD is shown as

sticks and the parts of NAD which are not solvent accessible is shown as lines. The visibility of NAD molecule shows that the binding pocket is solvent accessible. **E) Cartoon representation of Tm_AroKB NAD-binding pocket zoomed:** The panel shows the zoomed cartoon representation of Tm_AroKB NAD-binding pocket. The cartoon representation is shown in gray, while the nitrogen atom of Arg195 and oxygen atom of Asp316 forming the salt bridge is shown in blue and red and the salt bridge is shown in yellow. **F) Cartoon representation of An_AroKB zoomed:** The panel shows the zoomed cartoon representation of An_AroKB NAD binding pocket. An_AroKB contains an Asn46 and Pro184 instead of Asp and Arg in Tm_AroKB. The nitrogen of Asn is shown in blue while Pro oxygen is shown in red.

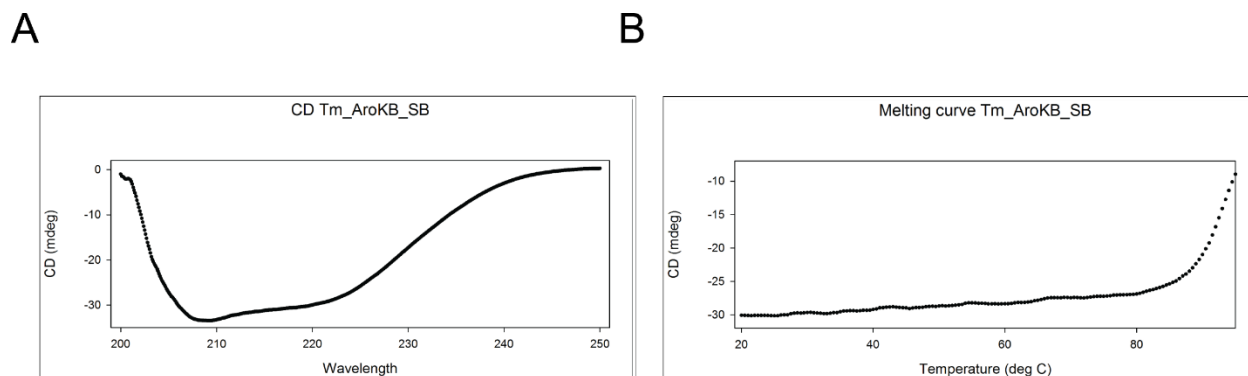


Figure 3.12 Thermal stability of Tm_AroKB_SB **A). CD profile:** 400 μ L of Tm_AroKB_SB at a concentration of \sim 0.3 mg/mL was measured in the CD spectrometer. The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarised light. **B) Thermal stability analysis:** 400 μ L of Tm_AroKB_SB at a concentration of \sim 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 85 $^{\circ}$ C.

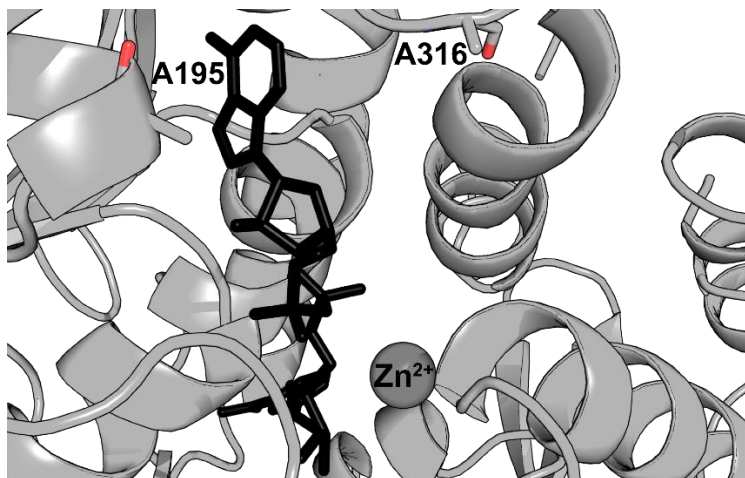


Figure 3.13 NAD-binding pocket of Tm_AroKB_SB. The figure shows the zoomed cartoon representation of one monomer of AroB of Tm_AroKB_SB co-crystallized with SKM and NAD. The cartoon is shown in gray color, NAD is shown in black and Zn^{2+} is shown in gray and is marked. The oxygen of Ala is shown in red. NAD is in the same orientation as seen in Tm_AroKB.

3.2.4 Crystal structure of the Tm_AroKB in complex with S3P and ADP

AroK catalyzes the conversion of SKM to S3P with the cofactors ATP and Mg^{2+} [223]. To follow the catalytic turnover in the crystalline state, the crystals of Tm_AroKB were soaked with SKM and ATP for time intervals ranging from 30 minutes to 24 hours. The crystals did not degrade during the soaking period and diffraction data could be collected at 2.75 Å. However, the structures, which were solved using the Tm_AroKB coordinates, always contained only SKM in complex with Tm_AroKB.

We hypothesized that the sulfate present in the ammonium sulfate of the crystallization condition can bind instead of phosphate of ATP at the p-loop, [133] and therefore, ATP might not be able to bind and catalyze the reaction. Thus, prior to these soaks, the crystals were soaked in 2.2 M sodium carbonate for thirty minutes to replace the sulfate present in the condition. Neither the prior soaking with sodium carbonate nor the later soaking with SKM and ADP led to crystal degradation, and the crystals diffracted to 2.75 Å. The structures were solved using the coordinates of Tm_AroKB and now contained both the products S3P and ADP with full occupancy (Figure 3.14). The conformation of the K domain is not affected particularly by the binding of either SKM or S3P. The binding of ADP shows that the position of the β -phosphates of ADP was nearly identical to the location occupied by the sulfate ion [133]. Only two previously reported AroK crystal structures (2IYZ, 3MUF) contain both the ligands S3P and ADP [133, 224]. The activity of AroK has been assessed in the crystalline state only in 2IYZ, while in 3MUF, the crystals were obtained together with S3P and ADP. The K domain from Tm_AroKB is the second crystal structure to report the catalytic activity in the crystalline state.

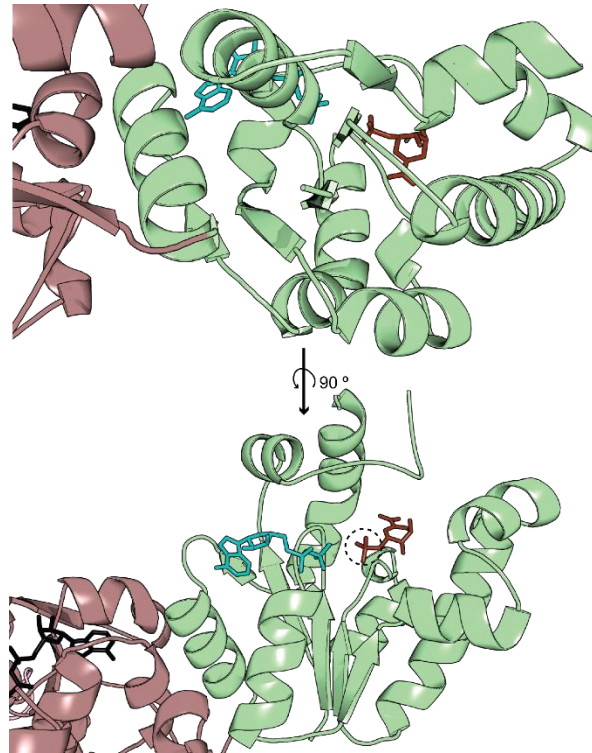


Figure 3.14 Tm_AroKB K catalytic activity in crystalline state: The Figure shows the K domain of one monomer of AroKB. Both the S3P and ADP are bound in the active site. β -phosphate of ADP is bound where a sulfate ion was bound in the previous structure.

3.3 AroEK

In this section, the bifunctional fusion protein AroEK, which catalyzes steps four and five of shikimate pathway is discussed (Figure 3.15). Unlike AroKB, AroEK catalyzes consecutive steps and therefore, this bifunctional enzyme could increase the flux through the shikimate pathway.

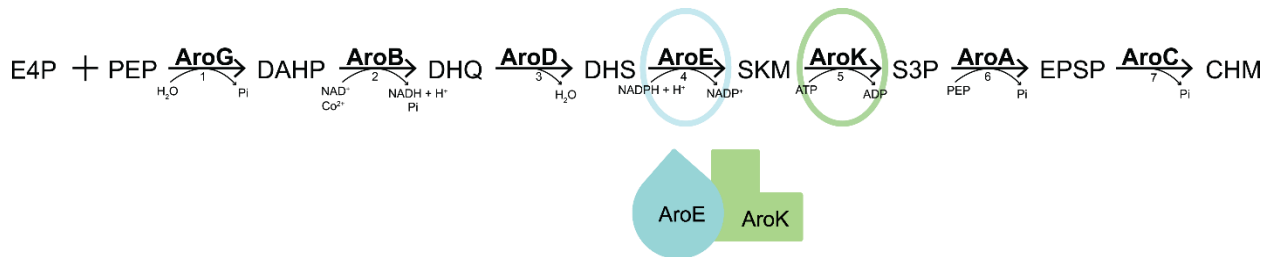


Figure 3.15 AroEK: The figure shows the hypothetical model of AroEK, which contains the fusion of the fourth and fifth steps of the shikimate pathway.

3.3.1 Biochemical and biophysical characterization

For experimental characterization, AroEK proteins from the organisms *Eubacterium bifforme* (Eb_AroEK) and *Faecalibacterium prausnitzii* (Fp_AroEK) were recombinantly expressed in *E.*

coli. AroEK from *E. biforme* gave an insoluble protein, which could not be solubilized with initial optimization (data not shown). Fp_AroEK also formed inclusion bodies when expressed in *E. coli* BL21DE3 at 37 °C but gave a soluble well-folded protein when expressed in BL21-Gold(DE3) at 25 °C. This bifunctional enzyme is annotated only as shikimate kinase in Uniprot.

F. prausnitzii is a dominant member of human intestinal microbiota therefore optimal growth temperature is 37 °C [225]. The biophysical characterization of Fp_AroEK shows a thermal melting temperature (T_m) of ~50 °C. It migrated as a single species in size exclusion chromatography with an average molecular mass of ~50 kDa as determined by static light scattering (Figure 3.16).

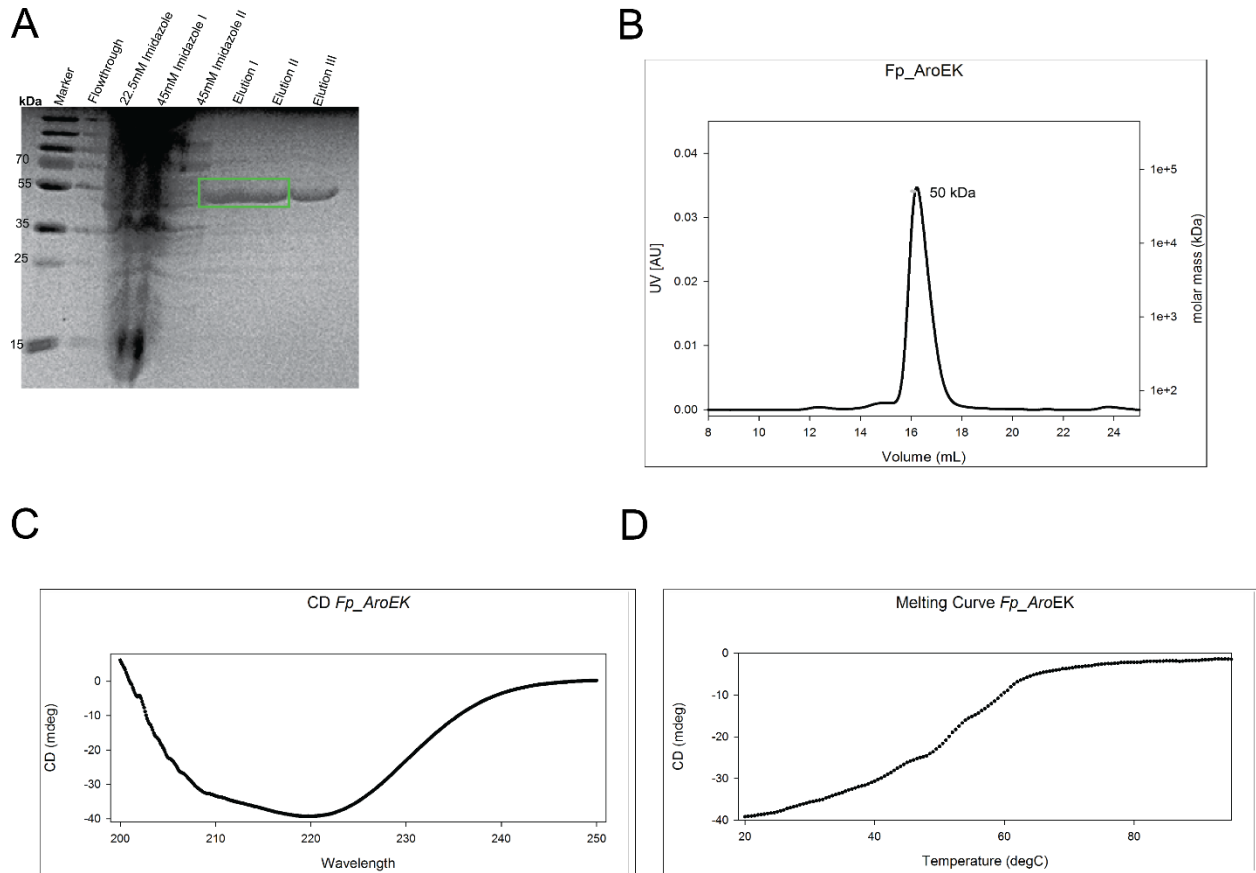


Figure 3.16 Biochemical and biophysical properties of Fp_AroEK: **A) SDS gel of FP_AroEK:** The gel shows the molecular weight of the monomer of the AroEK is 50 kDa. The gel shows 10 μ l of the samples loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B) Light scattering of FP_AroEK:** 100 μ l FP_AroEK was loaded at a concentration of 1 mg/ml (black profile), onto an S200 size exclusion column and the mass of the eluted particles (gray profile) in the peak area was analyzed with static light scattering. The horizontal axis of the curve shows volume in mL and the left axis shows the UV absorption. The molar mass on the right is shown in a logarithmic scale. **C) CD profile of FP_AroEK:** 400 μ l of FP_AroEK at a concentration of ~0.3 mg/mL was measured in the CD spectrometer.

The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarized light. **D) Thermal stability analysis of FP_AroEK:** 400 μL of FP_AroEK at a concentration of 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}\text{C}$ to 95 $^{\circ}\text{C}$. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 85 $^{\circ}\text{C}$.

3.3.2 Kinetic analysis of Fp_AroEK

The activity for the substrate ATP for the K domain was analyzed. The kinetic analysis of the K domain was conducted in the same way as discussed for Tm_AroKB in section 3.3.2. The kinetic analysis shows that Fp_AroEK follows Michaelis-Menten kinetics (Figure 3.16) and the kinetic activity falls in the same range as reported for K domains of *M. tuberculosis* [187] (Table 3.2). A detailed kinetic analysis and regulation has yet to be performed.

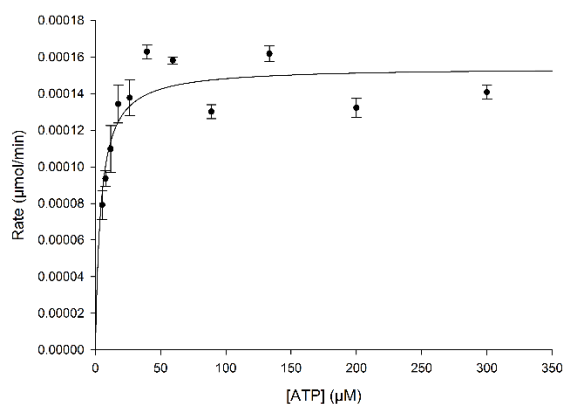


Figure 3.16 FP_AroEK K kinetics for ATP: The figure shows the kinetic analysis of K domain for ATP for Fp_AroEK. The enzyme concentration is around 10 nM.

Table 3.2 The kinetic parameters of Fp_AroEK

Enzyme	Sub (S) (μM)	$K_m \pm \% \text{ S.E.}$ (μM)	E_t (nM)	$k_{\text{cat}} \pm \% \text{ S.E.}$ (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \text{ M}^{-1}$)	R^2	Ref
Fp_AroEK	ATP	$4.17 \pm 19.6 \%$	14.9	$1.73\text{e}+0 \pm 3.07\%$	$4.14\text{e}+05$	0.649	TS

3.3.3 Crystal structure of Fp_AroEK in complex with SKM

Initially, Fp_AroEK crystallized in complex with SKM. The crystals diffracted to 2.7 \AA . The crystal structure of 4YOA was used as the search model for the K domain and the crystal structure of 5DZS was used as the search model for the E domain, for molecular replacement. The asymmetric unit contained one Fp_AroEK protomer, in which the constitutive domains are connected via a rigid interface. Again, the active sites are solvent-accessible (Figure 3.17).

SKM is bound to both E and K domains, which were part of the crystallization buffer. SKM can bind to both these sites, because it is the substrate for the K domain and the product of the E domain. The only difference between DHS (substrate of E domain) and SKM is the presence of hydrogen at C-3. Most of the reported AroE crystal structures contain SKM bound to the catalytic domain [33, 202, 226-228].

The domain structure of AroE is essentially identical to the previously reported instances. Some AroE contain two conserved *cis*-peptide prolines existing in a sharp turn linking a β -strand and the following α -helix [38, 119, 120]. Only one of such prolines Pro-55 is seen in the E domain of AroEK, which is the 2nd proline of the two. The structure of the K domain is similar to the K domain of AroKB reported in the previous section. The lid region is disordered in this crystal structure and therefore is not present in this structure.

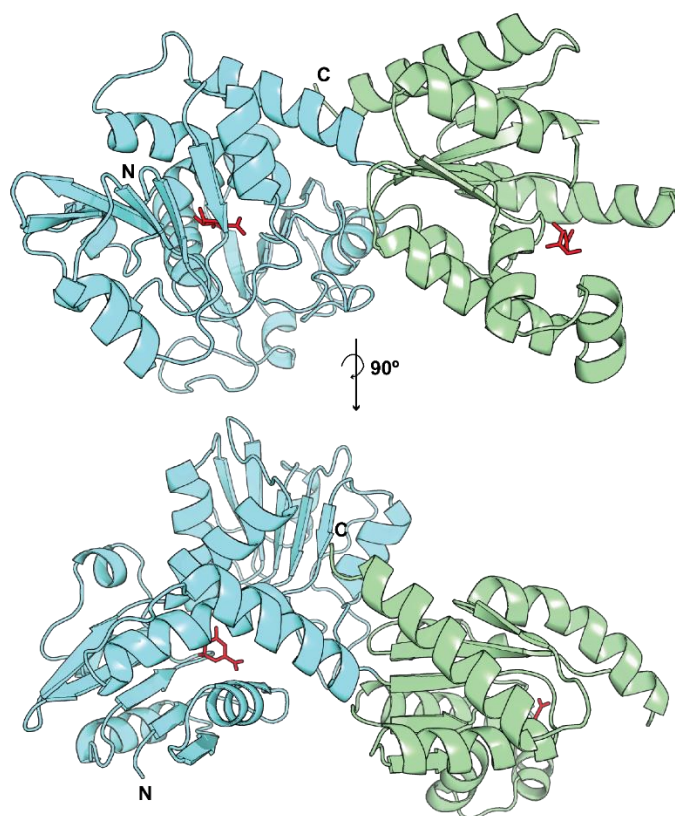


Figure 3.17 Crystal structure of Fp_AroEK in complex with SKM: The figure shows the crystal structure of AroEK. The N and C termini are marked. K domain is shown in green and the AroE domain is shown in cyan, SKM bound to both E and K domain is shown as red sticks.

3.4. Mb_AroKE

Like AroEK, AroKE also catalyzes steps four and five of shikimate pathway but the gene sequence indicates the presence of the K domain before the E domain. The difference between AroEK and AroKE is not only the different domain arrangement in the DNA sequence (Figure 3.18), but also the presence of a 30-40 aa linker between the two domains. Linkers separate multiple domains in a single protein; they are often rigid and function to prohibit unwanted interactions between the different domains [229]. To understand the difference between the two proteins, we aimed to determine the crystal structures of both proteins.

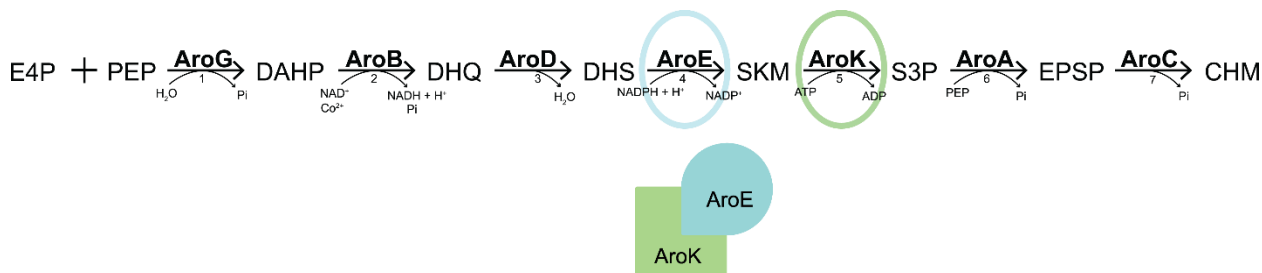


Figure 3.18 AroKE: The figure shows the hypothetical model of AroKE, which contains the fusion of the fourth and fifth steps of the shikimate pathway.

3.4.1 Biochemical and biophysical characterization

AroKE proteins were purified from *Methanoplanus petrolearius* (Mp_AroKE) and *Methanoregula boonei* (Mb_AroKE). Mp_AroKE was insoluble and formed inclusion bodies, which could only be purified in denatured state in the presence of either 6 M GnCl or 8 M Urea (Figure 3.19). Various attempts to refold the protein, including serial dialysis, on column refolding, and rapid dilution, failed.

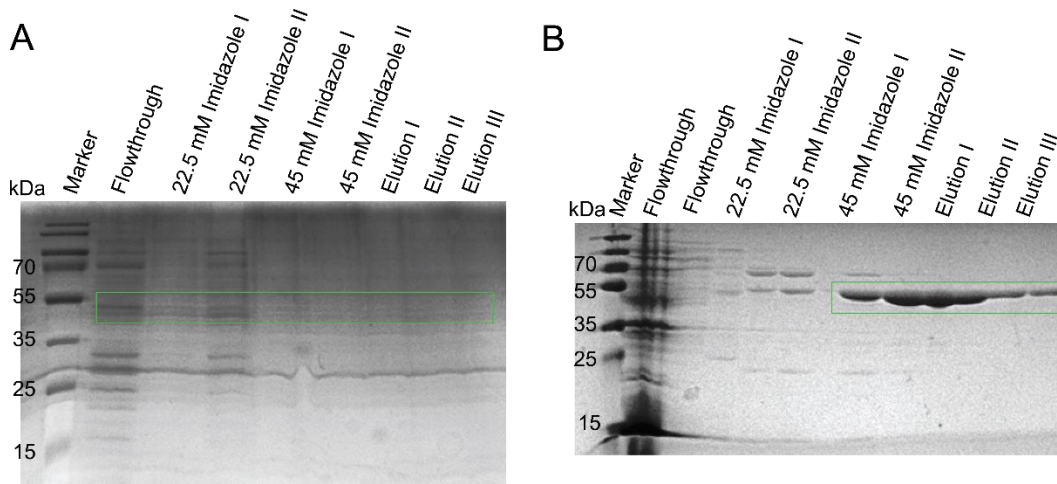


Figure 3.19 Presence of Mp_AroKE in inclusion bodies: A) SDS gel of Mp_AroKE after Nickel column in native conditions: The gel shows the complete absence of properly folded Mp_AroKE. The gel shows 10 μ l of the samples loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B) SDS gel of Mp_AroKE after Nickel column in denaturing conditions:** The gel shows that Mp_AroKE is expressed and can be purified in denaturing conditions. The gel shows 10 μ l of the samples loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 6 M GnCl, 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture.

Mb_AroKE also forms inclusion bodies when expressed in *E. coli* BL21DE3 cells at 37 °C but it is a well-folded protein when expressed in ArcticExpress cells at 10-12 °C, *M. boonei* is an acidiphilic, hydrogenotrophic methanogen. It was isolated from an acidic (pH 4.0-4.5) and ombrotrophic (rain-fed) bog located near Ithaca, NY, USA [230]. The purified protein has a thermal melting temperature (T_m) of ~50 °C as seen by CD. It migrated as a single species in size exclusion chromatography with an average molecular mass of ~51 kDa as determined by static light scattering (Figure 3.20).

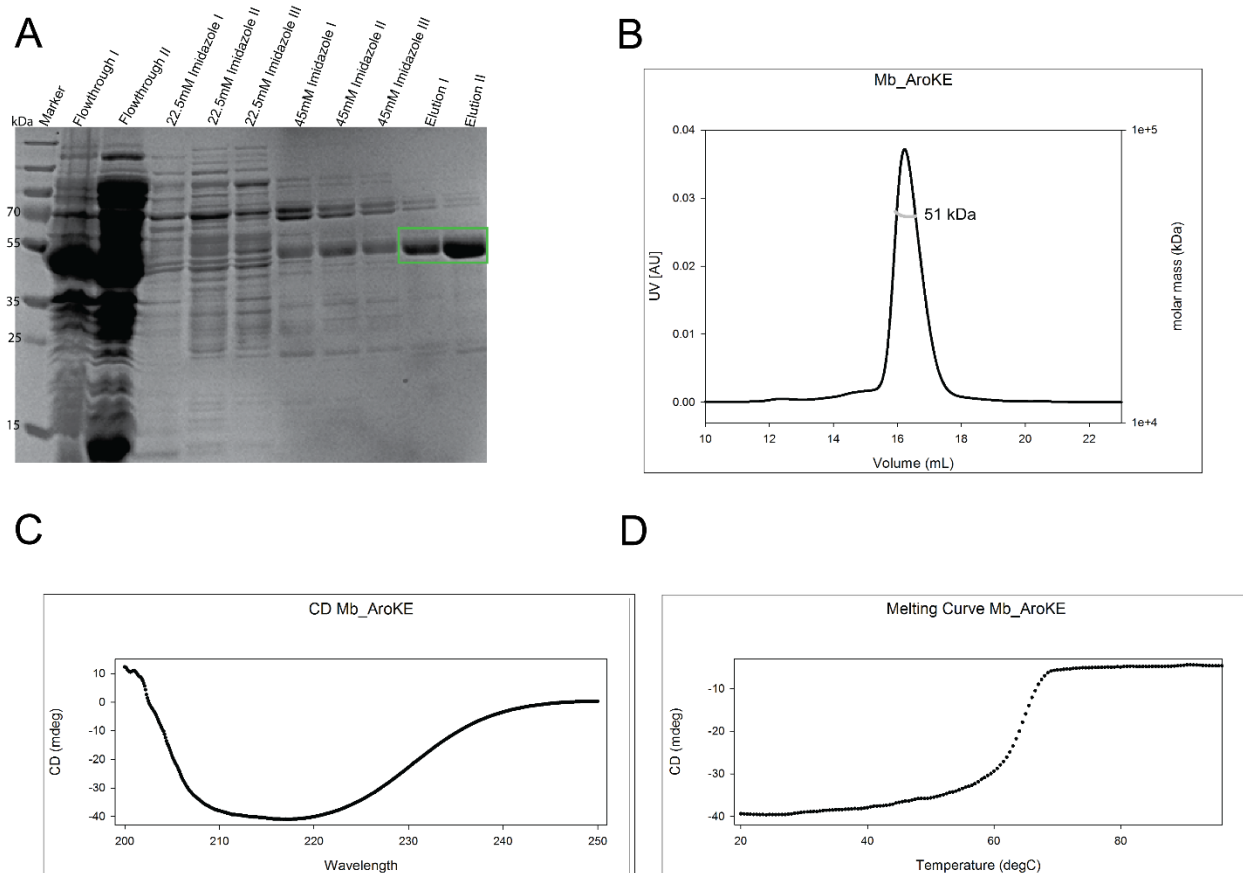


Figure 3.20 Biochemical and biophysical properties of Mb_AroKE A) SDS gel of Mb_AroKE: The gel shows the molecular weight of the monomer of the Mb_AroKE is 50 kDa. The gel shows 10 μ l of the samples

loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B). Light scattering of Mb_AroKE:** 100 μ l Mb_AroKE was loaded at a concentration of 1 mg/ml (black profile), onto an S200 size exclusion column and the mass of the eluted particles (gray profile) in the peak area was analyzed with static light scattering. The horizontal axis of the curve shows volume in mL and the left axis shows the UV absorption. The molar mass on the right is shown in the logarithmic scale. **C). CD profile of Mb_AroKE:** 400 μ l of Mb_AroKE at a concentration of \sim 0.3 mg/mL was measured in the. The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarized light. **D). Thermal stability analysis of Mb_AroKE:** 400 μ l of Mb_AroKE at a concentration of \sim 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 85 $^{\circ}$ C.

3.4.2 Kinetic analysis of Mb_AroKE

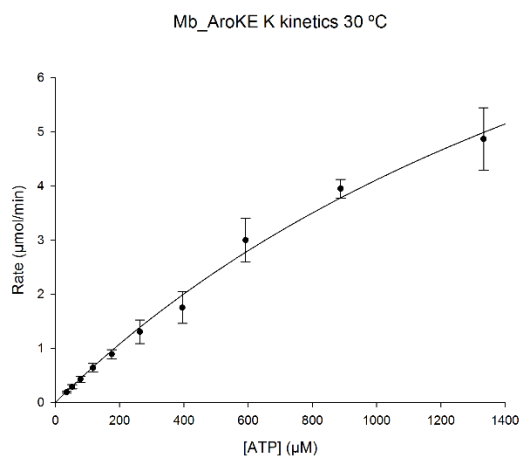


Figure 3.21 Mb_AroKE K kinetics for ATP: The Figure shows the kinetic analysis of the K domain for ATP for Mb_AroKE. The enzyme concentration is around 10 nM.

Initial kinetic assays showed an unexpected behavior. The kinetic activity for the substrate ATP of the K domain was assessed in the same way as for Tm_AroKB and Fp_AroEK in section 3.3.2. However, the kinetic analysis did not show a saturation even when an ATP concentration of 1.4 mM was tested. This behavior is different from what is observed for Tm_AroKB and Fp_AroEK. Further, the binding affinity measurements performed with MST failed to show the binding of SKM (data not shown).

A higher concentration to study the saturation has not been yet tested, but the requirement of a high amount of substrate is already known for *E.coli*, which has two shikimate kinases, AroL and AroK. Although AroL and AroK have greater than 30% sequence identity, their properties are very different. AroL has the K_M of 160 μ M for SKM and 200 μ M for ATP, while AroK has a K_M of 20 mM for SKM [244]. Both the chromosomal location and regulation are distinct for AroK and AroL. While AroL plays the dominant role in the shikimate pathway, AroK potentially plays a minor role in the pathway and might have other roles [130].

In addition to the bifunctional AroKE, *M. boonei* also has another monofunctional AroK, which is a 166 amino acid protein (Uniprot ID: A7I8L9) [245]. Interestingly, the bifunctional Mb_AroKE has only been annotated as AroE. A multiple sequence alignment between the K domains of Mb_AroKE, Mb_AroK, Ec_AroL, and Ec_AroK is shown in Figure 3.22. The figure shows that, all the conserved sequence elements for AroK are present in all four Mb_AroKE, Mb_AroK, Ec_AroL, and Ec_AroK, and they share more than 30% identity, which all functionally relevant residues conserved. Therefore, it could be that just like Ec_AroK, the K domain of Mb_AroKE plays only a secondary role in the shikimate pathway and has other functions.

```

Mb_AroKE      1 MK---RIVLFGYRGTGKTAIGTVLAQKLGVPFLDTDALVEQQAGRTIPEIFRDSGEAGFR
Mb_AroK       1 MK---NIILIGLPGAGKSTTGVI LAKTLGMGFIDTDLVQDRAGRILQEILDYEGPVAFLL
Ec_AroL       1 MT--QPLFLIGPRCCGKTVGMA LADSLNRRFVDTDQWLQSQ LNM TVAEIVEREEWAGFR
Ec_AroK       1 MAEKRNIFLVGPMGAGKSTIGRQLAQQLNMEFYDSDQEI EKRTGADVGVVFDLEGE EGFR

Mb_AroKE     58 AREREAVSGLPDR-DAI IATGGGVVMDPANMEHLRKESVCVLLSADPNVIGHRLAHAP--
Mb_AroK      58 TTEEKAIISLNCT-GT VIATGGSVVMSPKAI AHLKKTGVIVYLELSFAAMKRRLANIT--
Ec_AroL      59 ARETAALEAVTAP-ST VIATGGGIILTEFNRFHQNNGIVVYLCAPVSVLVNRLQAAP EE
Ec_AroK      61 DREEKVINELTEKQGI VLATGGGSVKSRETRNRLSARGVVVYLETTIEKQLARTQRDK--

Mb_AroKE    115 --RPALTSLSPT-DEITAMLKHRRPAYAAAADFCIDTGRTTAGEAAEKIILTLLGAGSIPD
Mb_AroK     115 --NRGIV-LLP-GQTLRHMFDRVPLYENYADLTVRC SKKDAESVVQEIVTGFQTG----
Ec_AroL     118 DL RPTLTGKPL-SEEVQEVLEERDALYREVAHIIIDATNEPSQ-VISEIRSALAQT----
Ec_AroK     119 -KRPLLHVETPPREVLEALANERNPLYEEIADVTIRTDDQSAKVVANQI IHML ESN----

```

Figure 3.22 Sequence alignment of the K domains of Mb_AroKE with Mb_AroK, Ec_AroL and Ec_AroK: The figure shows the sequence alignment of the K domains of the two SKs found in *M. boonei* with the two SKs found in *E. coli*. The Mg²⁺ coordinating aspartates are highlighted in yellow color, Walker A consensus sequence (Gly-XXXX-Gly-Lys-Thr/Ser; X is any amino acid) is highlighted in green color, and the Walker B consensus sequence (ZZ-Asp-XX-Gly; Z is a hydrophobic amino acid) is highlighted in cyan. The figure shows that although Mb_AroKE and Mb_AroK share more than 30% identity, they are not identical just like Ec_AroL and Ec_AroK. In addition, the Mb_AroKE is more similar to Ec_AroK than to Mb_AroK or Ec_AroL.

3.4.2. Crystal structure of Mb_AroKE in complex with SKM and phosphate

Initially, Mb_AroKE crystallized in a complex with SKM and phosphate. The crystals diffracted to 1.99 Å. The crystal structure of 4YOA was used as the search model for the AroK domain and the crystal structure of 2HK9 was used as the search model for the AroE domain, for molecular replacement. Again, the crystals contained one Mb_AroKE monomer in the asymmetric unit. Again, just like in the Tm_AroKB and Fp_AroKE structures, Mb_AroKE also has a rigid interface

connecting the two domains and all the active sites are solvent accessible. The additional linker forms an additional helical hairpin which is packed against the E domain. SKM is bound to AroE and a phosphate from the crystallization buffer is bound to the AroK domain. The structure of K domain is similar to the previously reported and discussed AroKs with the flexible lid disordered. Even though SKM was present in the crystallization buffer, it did not bind the K domain of Mb_AroKE. This further strengthens the hypothesis that the K domain of Mb_AroKE plays a secondary role in the shikimate pathway.

The structure of the E domain is similar to previously reported AroEs and contains the bound SKM in the catalytic domain (Figure 3.23). The two conserved cis-peptide prolines which connect β -sheet to α -helix are Pro-216 and Pro267.



Figure 3.23 Crystal structure of Mb_AroKE in complex with SKM: The figure shows the crystal structure of AroKE. The N and C termini are marked. The AroK domain is shown in green and the E domain is shown in cyan. The alpha helical linker between the two domains is shown in gray. SKM bound to AroE is shown as red sticks. Phosphate bound to AroK is shown as red and yellow sticks

Most interestingly, the observed domain arrangement in Mb_AroKE is entirely different from Fp_AroEK

3.6. *In silico* models of a hypothetical tetrafunctional AROM complex

Owing to the large size and low stability, the actual domain architecture of the AROM complex has been elusive since its discovery in 1960s. Therefore, using the respective orientations of four domains in four bifunctional enzymes as constraints, hypothetical tetrafunctional *in silico* models of the AROM complex were constructed, using the structures of Tm_AroKB, Fp_AroEK and Mb_AroKE. These enzymes together with At_AroDE constitute the four domains B, D, E, K of the AROM complex (Figure 3.24). To this end, the homologous domains of the bifunctional fusion enzymes were superimposed using secondary structure superimposition in COOT [183]. Two different models, one based on AroEK and another one based on AroKE were constructed.

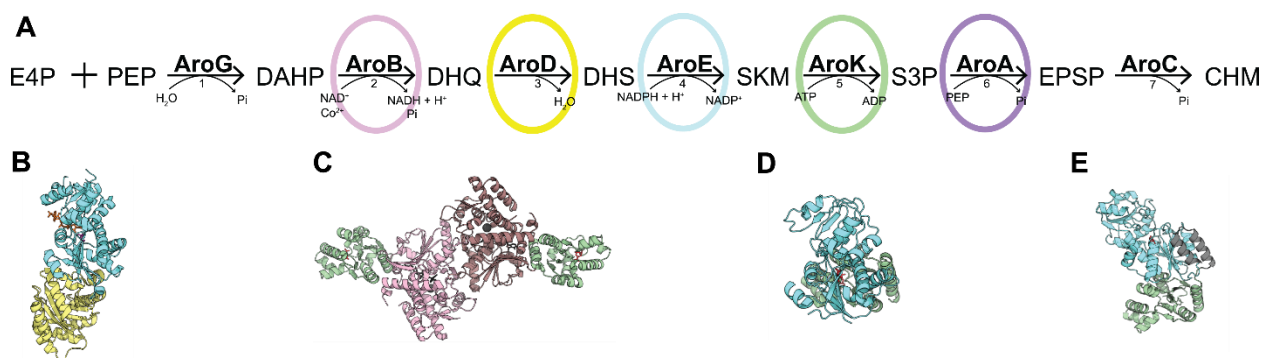


Figure 3.24. Pentafunctional AROM complex and the crystal structures of four bifunctional enzymes for the hypothetical tetrafunctional models: A) the five enzymatic activities of AROM complex, B) crystal structure of At_AroDE (PDB ID: 2O7S), C) crystal structure of Tm_AroKB (This study), D) crystal structure of bifunctional Fp_AroEK (This study), E) Crystal structure of bifunctional Mb_AroKE (This study)

To build the AroEK based model (Figure 3.25), the E domain of AroEK was superimposed on the E domain of AroDE (Figure 3.25.B). The resulting AroDEK showed no steric clashes. Then the K domain of AroKB was superimposed on the K domain of AroDEK and the resulting AroBDEK (Figure 3.25.C) showed no steric clashes. As AroKB exists as a dimer, the same procedure was repeated for the second monomer of AroKB (Figure 3.25.D). The complete model is also shown in surface representation in Figure 3.25E.

Mb_AroKE and Fp_AroEK have different domain orders; therefore, both plausible hypothetical models using both these enzymes were tested. The workflow for both the models was essentially the same except for employing Mb_AroKE to construct the AroKE based model. (Figure 3.26). For the sake of simplicity, only the surface representation of the AroKE based model is shown.

Figures 3.25 and 3.26 show that both models exhibit no steric clashes. The substrate path from B to D to E to K domains, highlighted in the figures, shows that both models bring the active sites closer to each other. This can potentially help in substrate channeling and it can avoid the dilution of reactants and facilitate the reactions. In addition, all the catalytic sites are solvent accessible and no steric hindrance is observed. Figure 3.27 compares both models and shows that both models have different domain arrangements. Thus, both these models could potentially resemble A) transient complexes of the individual enzymes of the pathway and B) an incomplete model of the AROM complex.

To test for the existence of such transient assemblies, pull-down experiments using the bifunctional proteins could be made. In a preliminary pull-down experiment with the recombinantly expressed Tm_AroKB protein (data not shown), no interacting enzymes of shikimate pathway could be identified. However, such a study should be repeated in the native environments of these enzymes, which was beyond the scope of this study.

With the two different and both possible hypothetical incomplete models of the AROM complex, a study of the AROM complex itself was needed to confirm these models. Therefore, to validate these models, studies on the AROM complex were initiated.

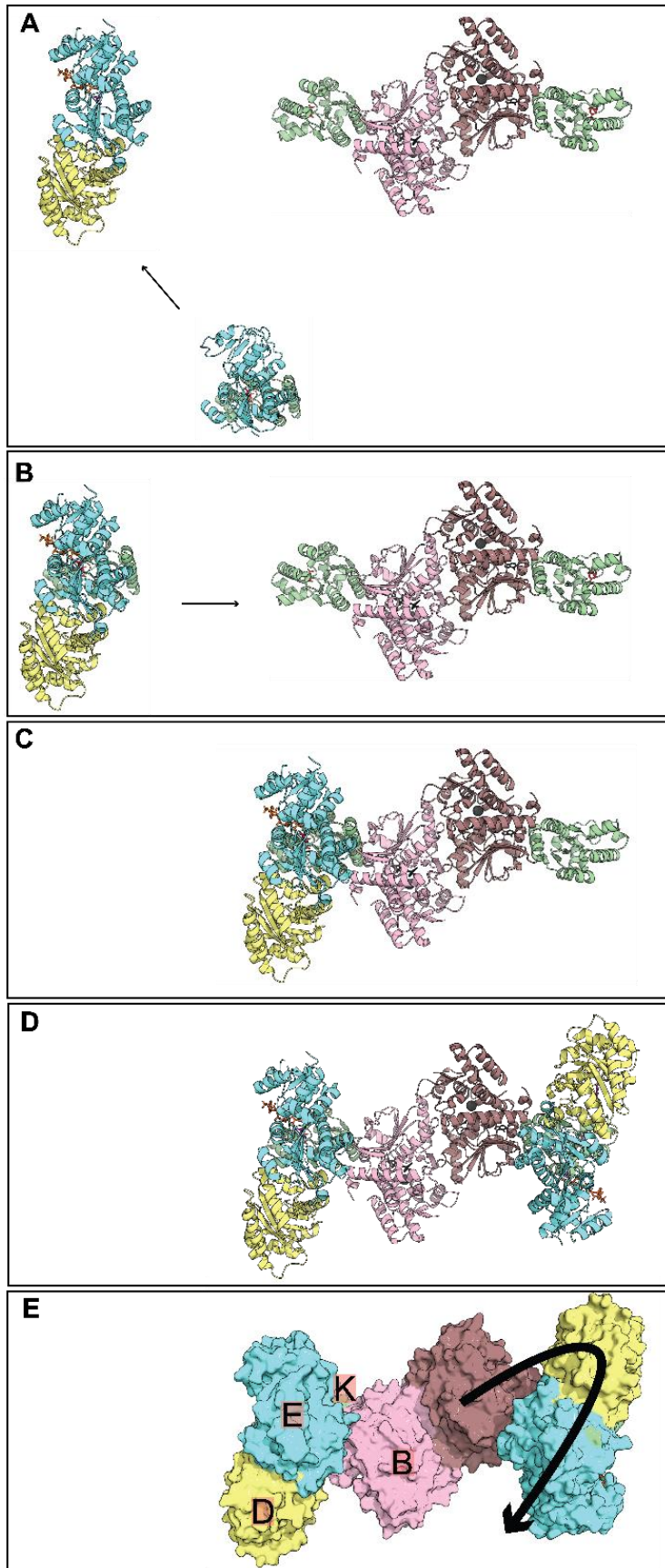


Figure 3.25. AroEK based hypothetical tetrafunctional model: The figure traces the steps in construction of the AroEK based model **A). Cartoon representations of the three bifunctional enzymes:** This panel shows the cartoon representations of At_AroDE, Fp_AroEK and Tm_AroKB. The two monomers of the AroB dimer are colored in pink and light brown to highlight the dimeric interface. **B). Cartoon representation of AroDEK and AroKB:** This panel shows the cartoon representation of the result of superimposing E domain of Fp_AroEK on the E domain of AroDE. **C) Cartoon representation of AroBDEK monomer:** This panel shows the cartoon representation of superimposition of K domain of Tm_AroKB on K on AroDEK. **D) Cartoon representation of AroEK based model:** The panel shows the dimer of AroEK based model. It was constructed by repeating the steps above for the second monomer of Tm_AroKB. **E). AroEK based model surface representation:** The surface representation of AroEK based model is shown in this panel. The line of symmetry divides the molecule into two halves; the individual domains are marked on one part of the dimer, while the other half traces the possible path that the substrate could follow to carry out the reactions.

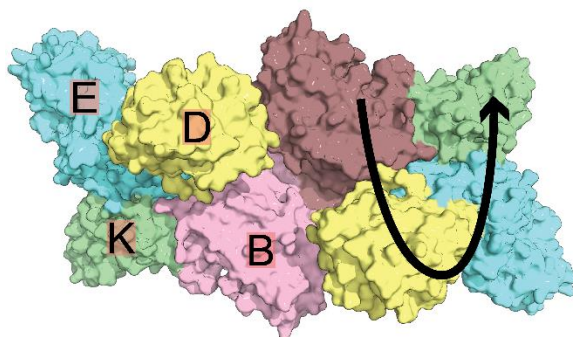


Figure 3.26 Surface representation of AroKE based model: The figure shows the surface representation of KE based model. The line of symmetry divides the molecule into two halves; the two monomers of the AroB dimer are colored differently. The individual domains are marked on one part of the dimer, while the other half traces the possible path that the substrate could follow to carry out the reactions.

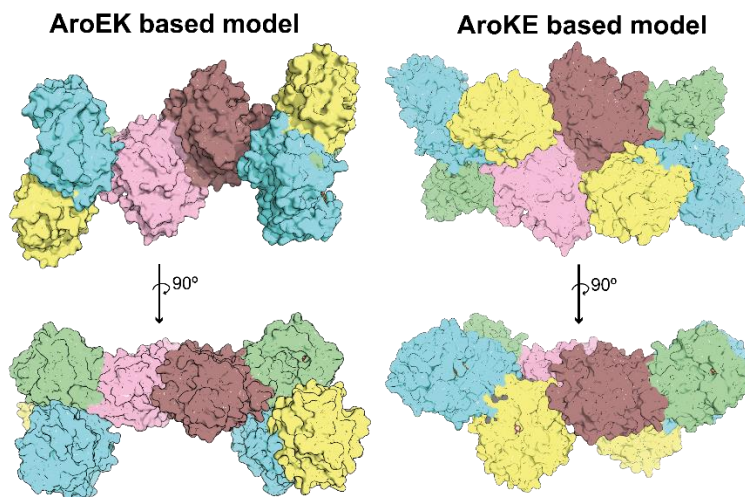


Figure 3.27 Comparison of AroEK and AroKE based models: The figure shows the surface representations of EK and KE based models. The different domain arrangements are highlighted.

3.7. AROM complex

Owing to its large size and fragile nature, a stable and homogenous AROM complex, which could be crystallized, could not be purified earlier [78, 79, 159-162]. Therefore, we searched for a thermostable AROM complex. The proteins from *C. thermophilum* are known to be very robust [231] [246]. Thus, the thermophilic fungi *Chaetomium thermophilum* and *Thielavia heterothallica* were selected for purifying a thermostable AROM complex. *C. thermophilum* is a thermophilic filamentous fungus. It grows on dung or compost with an optimal growth temperature of 50–55 °C [232]. *T. heterothallica* is the teleomorph to *Myceliophthora thermophila* [233]; it grows optimally at 38-45 °C [234].

3.7.1 Biochemical and biophysical characterization

The AROM complex was recombinantly purified from *C. thermophilum* and *T. heterothallica*. Both purified proteins are well-folded with a thermal melting temperature (T_m) of 65 °C as seen in CD. Ct_ARAM migrates as a single species in size exclusion chromatography with an average molecular mass of ~344 kDa as determined by static light scattering. (Figure 3.28). Ct_ARAM was visualized on an EM grid using negative staining. Negative stain EM can be used for visualizing homogeneity and/or heterogeneity of purified proteins sample [250]. As seen in figure (3.29), Ct_ARAM shows a homogenous sample with well defined particles and shows no instances of aggregation. It is readily visible that Ct_ARAM exists as a symmetric dimer and that the sample would be well suited for cryoEM.

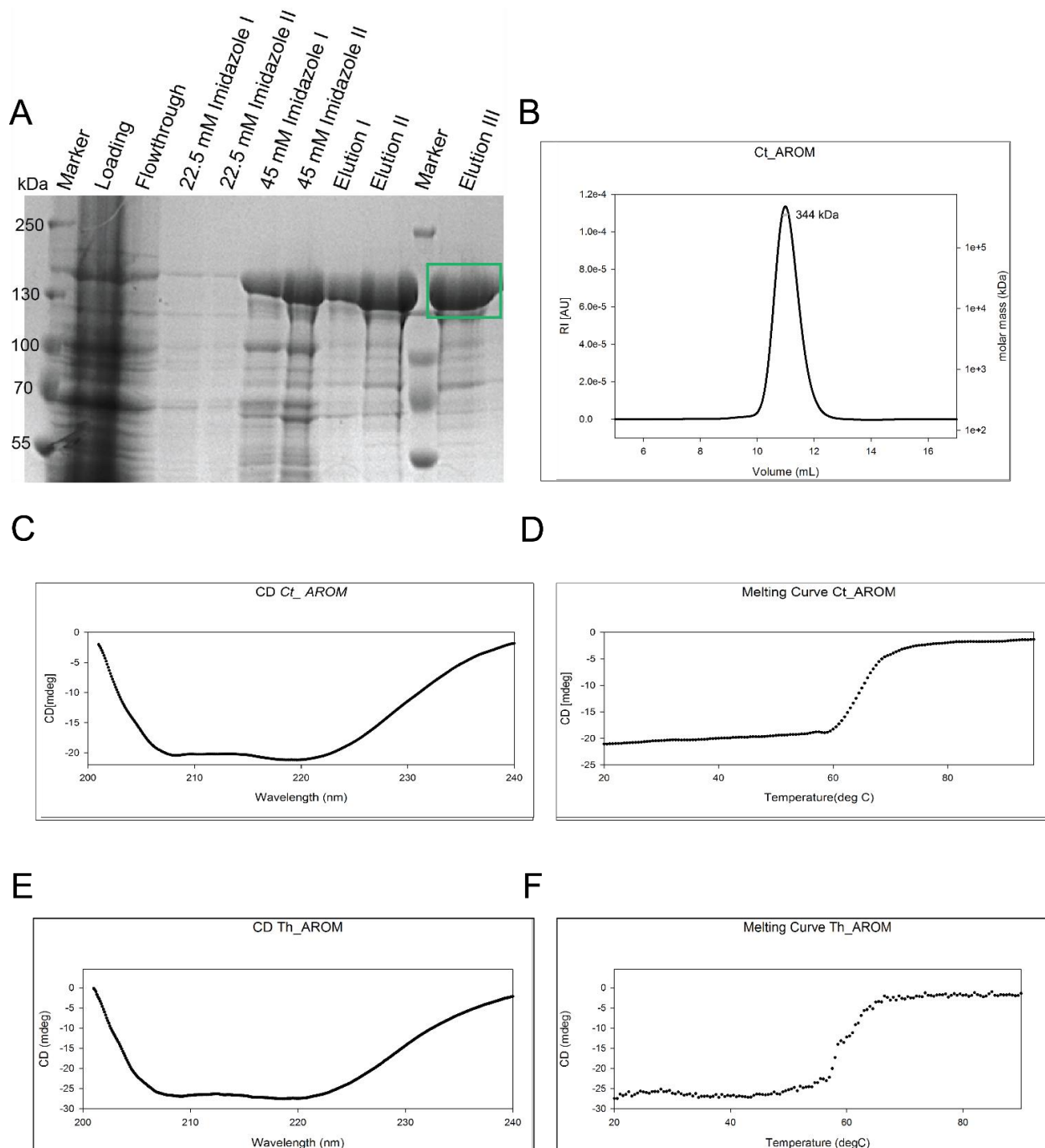


Figure 3.28 Biochemical and biophysical properties of Ct_AROM: **A). SDS gel of Ct_AROM:** The gel shows the molecular weight of the monomer of Ct_AROM is 172 kDa. The gel shows 10 μ l of the samples loaded after the purification from the Nickel column. The samples were loaded in the buffer containing 50 mM Tris, 300 mM NaCl and the respective imidazole concentration mentioned in the picture. **B). Light scattering of Ct_AROM:** 100 μ l Ct_AROM was loaded at a concentration of 1 mg/ml (black profile), onto a S200 size exclusion column and the mass of the eluted particles (gray profile) in the peak area was analyzed with static light scattering. The horizontal axis of the curve shows volume in mL and the left axis shows the UV absorption. The molar mass on the right is shown in a logarithmic scale. **C). CD profile of Ct_AROM:** 400 μ l of Ct_AROM at a concentration of \sim 0.3 mg/mL was measured. The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarized light. **D). Thermal**

stability analysis of Ct_ AROM: 400 μ L of Ct_ AROM at a concentration of \sim 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 65 $^{\circ}$ C. **E). CD profile of Th_ AROM** 400 μ L of Th_ AROM at a concentration of \sim 0.3 mg/mL was measured. The figure shows the far UV CD spectrum, plotted in mdeg versus wavelength in nanometers (nm) of polarized light. **F). Thermal stability analysis of Th_ AROM:** 400 μ L of Ct_ AROM at a concentration of \sim 0.3 mg/mL was measured at a temperature range from 20 $^{\circ}$ C to 95 $^{\circ}$ C. The temperature is displayed on horizontal axis while CD (mdeg) is shown on the vertical axis. The protein is stable up to 65 $^{\circ}$ C.

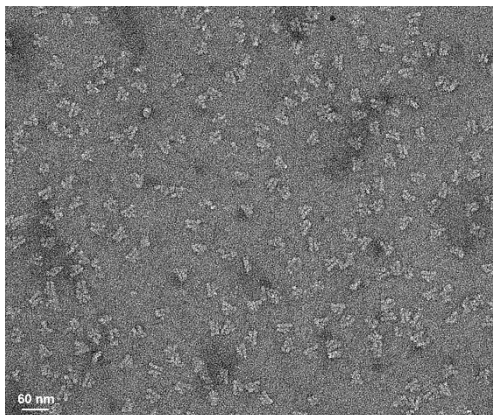


Figure 3.29 Negative staining of Ct_ AROM: The Figure shows, 0.1 mg/ml protein solution, stained with 1% uranyl acetate and examined with transmission EM. Homogenous, isolated particles are visible in the image.

3.7.2 Kinetic studies of Ct_ AROM

The catalytic activities of individual domains of AROM have been characterized from *N. crassa* and *A. niger* [77, 204, 216, 235]. Welch *et al* have also studied the effect of prior incubation of *N. crassa* AROM with DAHP without any other substrates, and they found an increase in the catalytic efficiencies of four of the five enzymes [77]. In addition, the catalytic activities of the individual domains have been compared from *N. crassa* and *A. nidulans*, [204, 216, 235]. The D domain is sensitive to presence of certain metal ions. Therefore, the sensitivity of D domain of *N. crassa* has been compared to *E. coli* [94]. Additionally, B, the N terminal domain of AROM, has been studied individually both structurally and kinetically from *A. nidulans* [61, 91, 216]. Further, the bifunctional BA domain from *A. nidulans* AROM has been expressed in *E. coli* and could complement AroB⁻ and AroA⁻ *E.coli* strains. However, when the monofunctional A domain is purified, it is not active, even when the B domain is present in trans [236].

As the studies on the individual enzymes have been carried out multiple times earlier, we wanted to study the catalytic activity of the whole complex. The AROM complex contains the activities of five enzymes (BDEKA) of the shikimate pathway (Figure 3.30). It catalyzes the formation of

EPSP from DAHP in the presence of NAD, $\text{Co}^{2+}/\text{Zn}^{2+}$, NADPH, ATP and PEP. Formation of EPSP cannot be monitored by spectrophotometric or fluorometric assay. The catalytic activity of A domain could be measured in four different ways. In the reverse direction, coupling PEP to A.) the PK and LDH reactions and monitoring the oxidation at 340 nm as mentioned for the K domain kinetics in section 3.2.2. [185, 237] B.) to the PK, HK and G6PDH reactions and monitoring NADPH fluorescence at excitation/emission wavelengths of 340/460 nm. In forward direction, C.) coupling to the CHM (AroC at 275nm) and/or anthranilate synthases, (TrpGD) in a continuous fluorimetric assay [238]. D.) Measurement of the released phosphate during the activity of A domain [237, 239].

Both methods A) and B) cannot be employed because the AROM complex because of interference with the activity of both the K and A domains. Therefore, the reverse activity of A domain cannot be studied as K domain releases ADP. Assay of A domain coupled to AroC and TrpGD is still being optimized in the laboratory. Additionally, not only A domain releases a phosphate but also B domain. Therefore, a phosphate release assay cannot be employed to study the entire AROM complex.

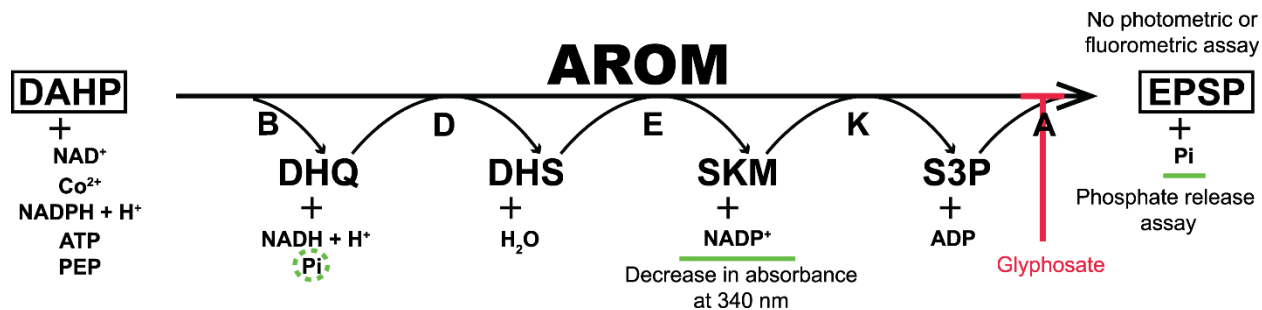


Figure 3.30 Kinetic assay of AROM complex: The Figure shows that AROM complex alone catalyzes the five activities of shikimate pathway. It is also highlighted that glyphosate inhibits the activity of A domain of the AROM complex.

The assay for the K domain, mentioned in the section 3.2.2 cannot be utilized to study the BDEK activities because the NADPH consumption by the AroE activity will compete with the NADPH production from the coupled PK/LDH assay. Therefore, the assay of AROM complex was divided into two parts. The activities of AroBDE were studied together, and the activities of AroKA were studied separately. The data analysis is as explained in section 2.4.2.3 (for Tm_AroKB kinetics).

3.7.2.1 Kinetic analysis for BDE domains

The activities of the three domains B, D and E were measured by studying activity of E domain. The activity of E domain can be analyzed by monitoring the formation of NADP⁺ in reverse direction [240] and in forward direction monitoring oxidation of NADPH at 340 nm [194]. Here the catalytic activity of E domain was monitored in the forward direction. AroB requires a divalent cation such as Co²⁺ or Zn²⁺ for its activity. The catalytic activities were compared in the presence of both Zn²⁺ and Co²⁺. The reaction rate was faster in the presence of Co²⁺ (data not shown). Table 3.3 lists all the substrates required for the AROM complex. The substrates until (including) the E and substrates until the K domain of the AROM complex are also listed. For initial studies, the reaction was carried out in the presence of all the substrates (Figure 3.31A). Ct_ AROM is stable up to 65 °C. Therefore, the kinetics for BDE domains was studied at 30 °C, 40 °C, 45 °C and 50 °C. The reaction cocktail containing all the components was incubated at the desired temperature for 15 minutes before adding the enzyme (10 nM) to start the reaction.

To test for potential product/substrate inhibition effects, the reaction was also performed with the substrates until the E domain and with substrates until the K domain. However, the first results of the kinetic analysis of substrates until K were noisy and the data could not be analyzed. This part of the study is yet to be repeated.

Table 3.3 Substrates for AROM complex kinetics

All substrates	Substrates until E domain	Substrates until K domain
DAHP, MgCl ₂ , CoCl ₂ , NAD, NADPH, ATP, PEP.	DAHP, MgCl ₂ , CoCl ₂ , NAD, NADPH.	DAHP, MgCl ₂ , CoCl ₂ , NAD, NADPH, ATP.

The activity of AROM complex increases with an increase in temperature until 45 °C (Figure 3.31A). In addition, the Ct_ AROM complex exhibits product based inhibition. The graph with substrates until E shows that the activity decreases substantially, in absence of the complete substrate cocktail. The data at 50 °C was very noisy and could therefore not be analyzed fully. However, the maximum velocity at 50 °C was similar to that at 45 °C, so the reaction was not studied at higher temperatures.

3.7.2.2 Kinetic analysis of *E. coli* enzymes

For a comparison between the monofunctional enzymes in *E. coli* to the pentafunctional AROM complex the activities of *E. coli* AroB, AroD, AroE, AroL and AroA were measured at AroE as

outlined for AROM complex above. The enzyme cocktail contained the five *E. coli* enzymes at a concentration of 30 nM.

To compare the catalytic activities of the monofunctional *E. coli* and the pentafunctional AROM complex, the assay was carried out in the same way as that of BDE kinetics of Ct_AROM. The enzyme concentration employed was three times that of the AROM complex, otherwise the reaction could not be measured optimally (Figure 3.31B). As can be seen in Table 3.4, the k_{cat}/K_M value of *E. coli* enzymes is three (2.9) times less than that of the AROM complex. Therefore, the pentafunctional AROM complex containing five different activities on one chain is almost three times faster than a mixture of the monofunctional *E. coli* enzymes. However, as *E. coli* and *C. thermophilum* are completely different organisms, further studies are needed to assess the potential benefits of the AROM complex as compared to monofunctional enzymes. Studies comparing the pentafunctional AROM complex to five AROM mutants, each lacking four activities could indicate if and to which extent the presence of all the activities on the same chain aids the catalytic activity.

3.7.2.3 Kinetic analysis of KA domains

To study the remaining two activities of the AROM complex, the activities of the KA domains was studied together. As described in the section 3.7.2 we could only study AroA activity by measuring the released phosphate. To measure the released phosphate continuously, the Enzcheck phosphate release assay kit [188] was employed. The activity of K and A domains was analyzed starting with SKM. The linear region for the Enzcheck phosphate assay extends until 150 μM , which was used as the maximum substrate concentration of this assay. Further, MESG is heat sensitive, the assay could only be performed at 30 °C. Figure 3.31 shows the KA kinetics of the AROM complex.

Further, the effect of glyphosate on the AROM complex A was studied. At 150 μM S3P concentration, the reaction was carried with 200 μM of PEP, with and without 200 μM of GLP (Figure 3.31D). Although, only the velocity at a single substrate concentration of 150 μM was determined, the graph clearly shows a reduction of at least 25% in presence of GLP, confirming that GLP acts as a competitive inhibitor of the Ct_AROM AroA moiety [140].

As seen in the crystal structure of Ct_AROM, (Figure 3.32) the B and A domains are next to each other. Therefore, it was hypothesized that this arrangement aids in regulation, i.e., A domain (last catalytic subunit) might be regulating B, (first catalytic subunit). Thus, the BDE kinetics were repeated in the presence of GLP. However, no influence of GLP on the kinetic behavior was observed, so no regulatory effects between the domains could be determined. This suggests that the domain arrangement of the AROM complex is optimized to facilitate substrate concentration rather than regulation. Further experiments are needed to validate this. Table 3.4 summarizes the kinetic analysis of Ct_AROM.

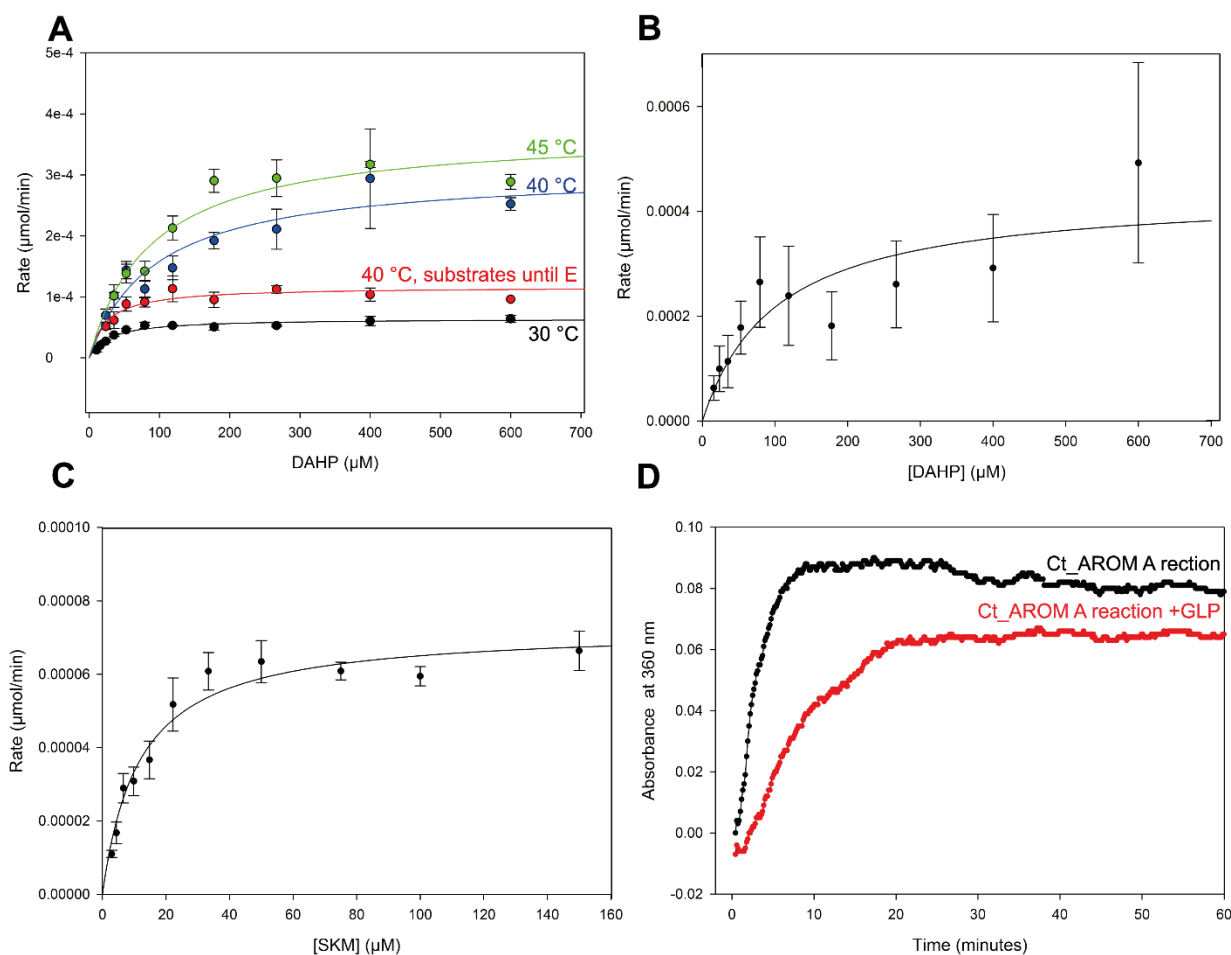


Figure 3.31 Ct_AROM kinetics: The figure shows the kinetic analysis of the AROM complex. **A). BDE kinetics:** This panel compares kinetics of BDE domains at different temperatures, and different colors represent different experiments. The activity increases with an increase in temperature. The red color shows the experiment conducted in the absence of ATP and PEP. The activity when not all the substrates are present is significantly decreased **B). BDE kinetics with the *E. coli* enzymes:** The experiment shows the kinetics carried out by AroB, AroD, AroE, AroL, AroA. The experiment was carried out at an enzymatic

concentration of 30 nM. **C). KA kinetics:** The figure shows Ct_ AROM KA kinetics at 30 °C. **D). Ct_ AROM glyphosate inhibition:** The panel shows the absorbance due to phosphate release.

Table 3.4 Catalytic activities of AROM complex

Activity	T (°C)	K _M ± % S.E. (μM)	E _t (nM)	k _{cat} ± % S.E. (s ⁻¹)	k _{cat} /K _M (s ⁻¹ M ⁻¹)	R ²
BDE for AROM (all sub)	30	28.14 ± 16.5%	10	4.27 ± 4.06%	1.52e+5	0.831
BDE for AROM (all sub)	40	94.33 ± 34.5%	10	20.49 ± 11.8%	2.17e+5	0.600
BDE for AROM (sub until E)	40	23.96 ± 33.6%	10	7.88 ± 7.22%	3.29e+5	0.495
BDE for AROM (all sub)	45	87.90 ± 20.36%	10	24.78 ± 6.5%	2.82e+5	0.825
BDE <i>E.coli</i> (all sub)	30	101.23± 64.8%	30	9.72 ± 25%	9.60e+4	0.334
KA AROM (sub from K)	30	11.68 ± 16.69%	10	4.84 ± 4.6%	4.14e+5	0.862

The kinetic study here analyzes the catalytic parameters of various enzymatic activities together, therefore the error rates are observed to be higher. Further, the reaction volumes for BDE kinetics were 25 μL, such a low volume can also contribute to the error rates. Higher error rates were also observed for Tm_AroKB, when the reaction was conducted with 50 μL volumes.

3.7.3 XL-MS studies Ct_ AROM

In crosslinking mass spectrometry (XL-MS), proteins are chemically crosslinked in their native or native-like state and then cleaved to peptides by enzymatic digestion to identify the sequence of crosslinked peptides with tandem MS [241]. The DSS crosslinker was employed in the AROM crosslinking experiments. DSS is a six-carbon succinimide-type conventional lysine cross-linking reagent [242]. DSS, when fully extended, has a 11.4 Å long linker arm, allowing C_α atoms of crosslinked lysine residues to be up to ~24 Å apart. Adding a tolerance of ~3 Å to both sides, DSS can crosslink lysine residues with distances up to 26–30 Å between C_α atoms [243]. Crosslinking was carried out at two different concentrations of DSS (250 μM and 1 mM and respectively) to confirm no occurrence of aggregation tendencies with stronger cross-linking. The data looks rather similar for both cross-linker concentrations. A large number of cross-links are observed, between 85 and more than 100. The number of false positive crosslinks of the data sets is <5%. The structure of the AROM complex could not be predicted from the XL-MS data alone. However, the data already suggested that the A domain of the AROM complex was closer to the B domain of the AROM complex, rather than the proximity of the K and B domains as predicted by the AroEK and AroKE based models. Therefore, it indicated that models do not predict the organization of the AROM complex.

3.7.4. Crystal structure

Initially, crystallization trials were set up both for Ct_AROM and Th_AROM both in Apo state and in combination with various ligands. While testing different crystallization conditions it was observed that both the proteins only crystallized in the presence of a Hepes buffer and TCEP, rather than Tris and β -ME. Initial crystals diffracted to 8 Å. As Ct_AROM crystals proved to be better reproducible, further crystallization optimization was continued with Ct_AROM.

Ct_AROM only crystallized in the presence of SKM, DHQ and NAD or alternatively with DHS, DHQ and NAD. It was observed that DHS and SKM can be interchangeably used for the crystallization condition, but the crystals did not form in the absence of DHQ. The crystals of native Ct_AROM diffracted at a resolution of 3.1-3.3 Å. Due to the large size of the protein, the crystal structure of the native Ct_AROM could not be solved by molecular replacement. Attempts to soak the crystals with platinum, mercury and heavy metal salts for experimental phasing deteriorated the data quality. Therefore, crystals were reproduced using Se-met containing Ct_AROM, which enabled structure solution via SAD phasing.

The structure contains two molecules in the asymmetric unit, forming a functional dimer (Figure 3.32). It is the first reported crystal structure of the AROM complex. The dimerization is mediated by the B and D domains. In both monomers, the five constituent domains are connected via rigid interfaces and all the active sites are solvent accessible. However, it is visible that two halves of the crystal structure are not similar. The two monomers occupy slightly different conformations, potentially corresponding to different catalytic states of individual domains. The structure shows that the B domain is connected to the A domain rather than the K domain as predicted by the *in-silico* models. Therefore, both models based on AroKE and AroEK are not correct in predicting the domain arrangement of the AROM complex (Figure 3.33). When the crosslinks obtained through XL-MS were mapped on the AROM complex, it was evident that the crosslinks agree with the crystal structure to large extents (Figure 3.32 C).

The structure contains NAD and Zn^{2+} bound to both B subunits, DHQ or DHS bound to both D subunits, SKM bound to both E subunits, both K subunits and one A subunit. The electron density at this resolution cannot distinguish well between DHQ and DHS, and therefore it is not clear

whether DHS or DHQ is bound to the D domain. However, as the protein was co-crystallized with DHQ, it seems the more probable ligand.

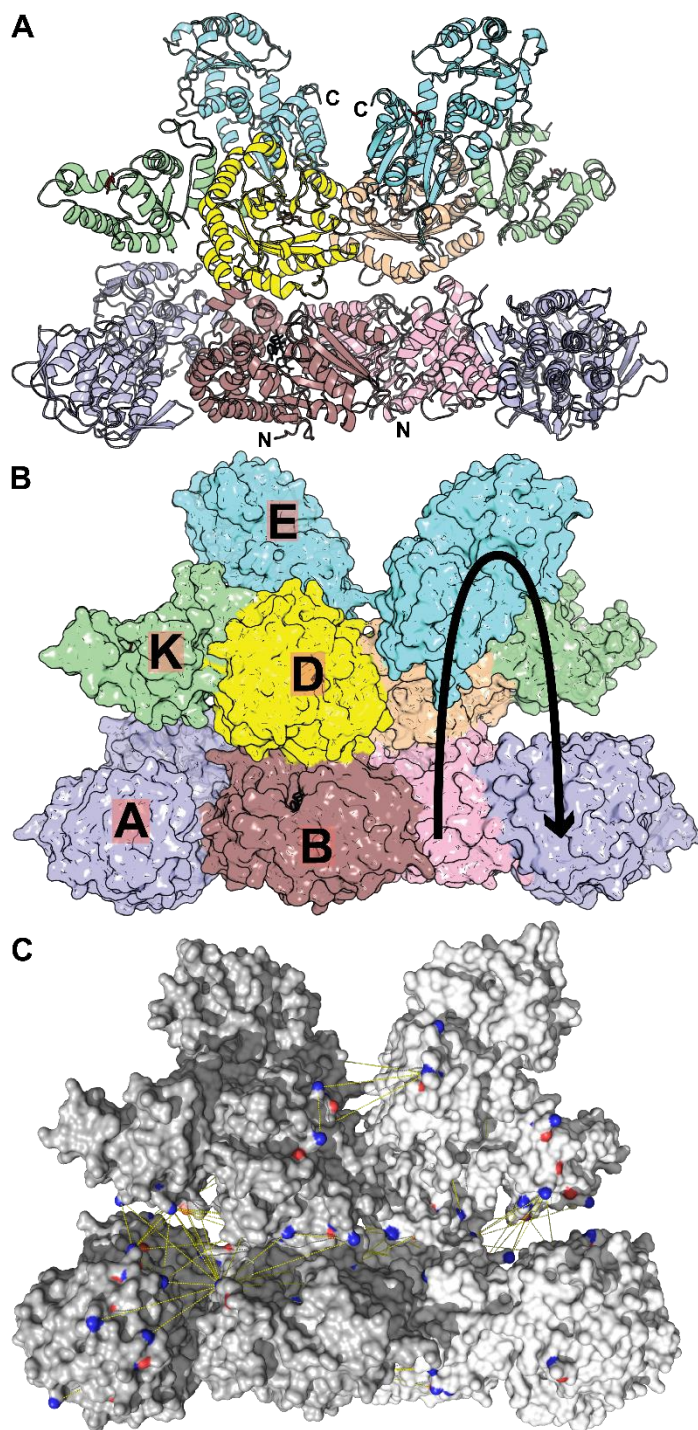


Figure 3.32 Crystal structure of Ct_ARAM: The figure illustrates the crystal structure of the pentafunctional AROM complex. **A). Cartoon representation:** The structure contains a homodimer. N and C termini are shown. SKM is bound to both E, both K and one A domain and is shown in red sticks. NAD

bound to B domain is shown in black sticks. DHS, bound to D domain is shown in brown sticks. The line of symmetry divides the structure in two monomers. **B). Surface representation:** The individual domains are marked; the B and D domains are colored differently to highlight the dimeric interface. The path, followed by the substrates, is traced in the structure. **C). XL-MS data mapped on the crystal structure of Ct_AROM:** The crystal structure is shown in gray with the dimer highlighted in two shades of gray. The crosslinks are shown in yellow and the residues connected with the crosslinks are shown in blue and red.

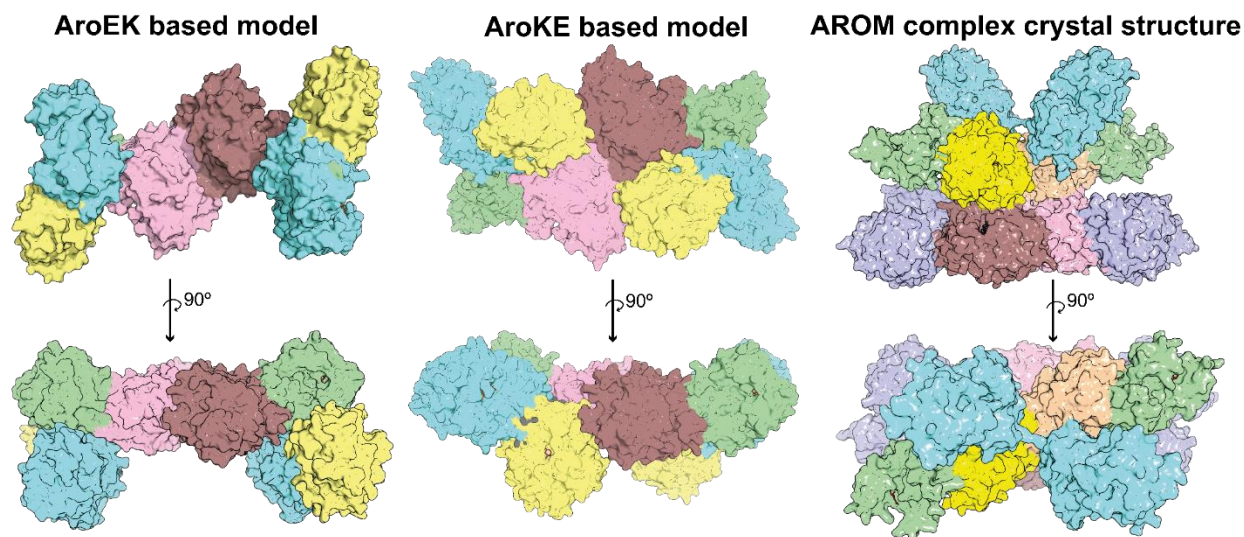


Figure 3.33 Hypothetical tetrafunctional AROM models vs crystal structure of AROM complex: The Figure shows that arrangement of the AROM complex is different from both predicted models.

The A domain on the right in Figure 3.32 contained additional electron density in the S3P binding site, which was best interpreted with a SKM molecule, while the A domain on the left did not contain ligands. Even though SKM is not the usual substrate of AroA, two other crystal structures are reported with bound SKM (PDB ID: 2AA9 [244] and 3SLH [143]). SKM can bind to the active site in the same orientation as S3P, with the 5-OH pointing towards the PEP-binding site [244].

3.7.5 Conformational space of Ct_AROM

The structure shows a compact assembly of the individual domains, in which each domain forms an interface to at least one other domain. Ct_AROM has both rigid and flexible parts and four of the five domains (B,E,K,A) can exist in different conformations depending on the catalytic state [92, 134, 200, 224, 228, 244, 245]. Therefore, the four domains B, A, K, and E, exhibit different domain movements based on their catalytic states.

The crystal structure shows only a static picture of these four domains. Therefore, the conformational space of Ct_AROM was explored using representative structures for each of the

four domains from the PDB. For the B domain (Figure 3.32, Figure 1.5), the Rossmann-fold is rigid at the dimer interface, while the α -helical sub-domain can undergo hinge motions. The attachment to α -helical B sub-domain confers additional flexibility to the A domain. Therefore, it can adopt various conformations depending on its own catalytic state and depending on the relative motion of the B domain. Such a cumulative effect is not found for the K and E domains, which are both attached to the rigid D domain, so they can have different conformations depending on only their own catalytic states.

The procedure outlining the mining of the representative structures is described in the section 2.7.8. The conformations of the crystal structures of the different monofunctional enzymes from the PDB in various catalytic states were mapped onto the Ct_AROM structure and thus, a structural ensemble of the accessible conformational states of the AROM complex was generated. Further, an analysis of this ensemble indicated that the individual domains can adopt essentially all conformations that occur throughout the individual catalytic cycles without steric clashes to other domains. To illustrate this conformational space (CS_AROM), a movie of AROM sampling the different conformations was generated, highlighting that the structure of the AROM complex is optimized for the spatial compatibility of its constitutive domains with respect to their conformational changes. The movie is accessible under the following link:

<https://owncloud.tuebingen.mpg.de/index.php/s/SznO2l0DrHoYEmB>.

Further, the XL-MS data was re-evaluated using CS_AROM. To study the distribution of the crosslinking restraints predicted by the XL-MS experiment graphically, each of the crosslinks was plotted against distance in Å. Similar crosslinks were computed from the different models generated in computing CS_AROM. Further, these crosslinks were plotted on the same graph and the possible crosslinks within the conformational space were compared to the experimental crosslinks (Figure 3.34). As seen in the Figure, most of the residues of Ct_AROM fall in the range below 30 Å, which suggests that AROM complex exists mostly in similar conformations in solution and in crystalline state. The distribution further indicates that several crosslinks correspond to conformations in the conformational space beyond the crystal structure, indicating that also other conformations are sampled in solution.

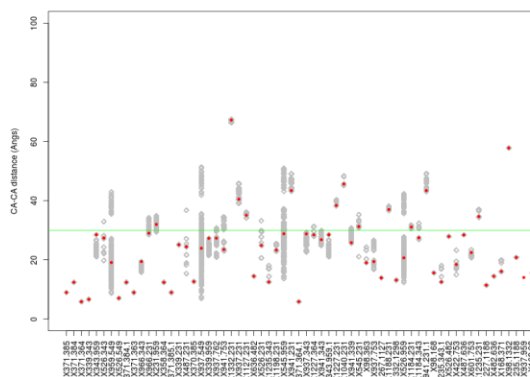


Figure 3.34. Graphical distribution of XLMS distance restraints: The figure shows the crosslinks between two lysines vs. the distance in Å. The experimental restraints obtained for the AROM complex are shown in red and the calculated distances for the models possible in the conformational space are shown in gray. Green line shows the distance of 30 Å. The concentration of DSS used for this experiment was 250 μ M.

3.7.6 SAXS experiments for Ct_AROM

Small-angle X-ray scattering (SAXS) studies the X-ray scattering of macromolecules in solution and predicts the overall shape and structural transitions of biological macromolecules in solution. It provides low-resolution information on the shape, conformation and assembly of proteins, and various other macromolecular complexes [246]. To understand the possible orientations represented by AROM complex, SAXS experiments were conducted both with Apo Ct_AROM and Th_AROM and in the presence of various ligands.

The initial data analysis for Ct_AROM apo protein carried out in Primus [195] is shown in Figure 3.35. The logarithmic plot of the scattering intensity I vs. q , (Figure 3.35 A.) shows the complete range of raw data after buffer subtraction. It does not indicate any aggregations or inter-particle interactions. This is further supported by the Guinier plot of $\log(I(q))$ against q^2 (Figure 3.35 B.), which yields a straight line [247].

The Kratky plot, $I(q) \cdot q^2$ vs. q (Figure 3.35 C.), exhibits a bell-shaped curve peak at low q and it converges to the q axis at high q . This indicates that Ct_AROM is a globular protein without pronounced flexibility. The Pair-wise distance distribution function $P(r)$, obtained by indirect Fourier transformation of the scattering curve exhibits a single peak and is thus further indicative of a globular proteins (Figure 3.35D). It further indicates a maximum length in the particle dimension, D_{max} , of 165.0 Å, as compared to about 180 Å in the crystal structure. The agreement

in the values of R_g and $I(0)$ calculated both from the $P(r)$ function and the Guinier plot indicate a good quality of data (Table 3.6)..

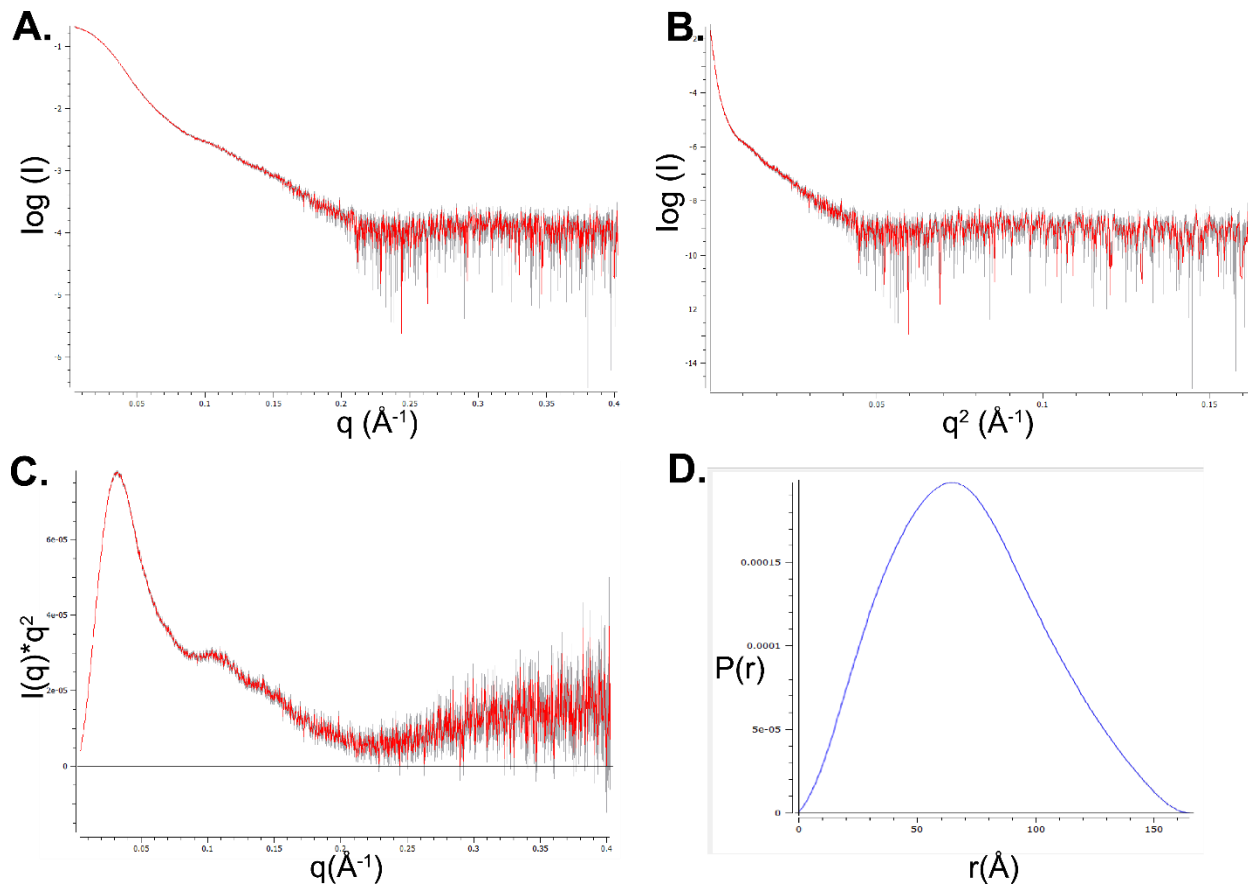


Figure 3.35 Plots estimating quality analysis of SAXS data **A). Logarithmic plot of the scattering intensity I (s) (in arbitrary units) vs. s (in \AA^{-1}):** This plot shows the complete range of the raw data after buffer subtraction **B) Guinier analysis:** The Guinier analysis shows the logarithmic scale of intensity vs s^2 which is a straight line and thus shows no aggregation. **C). Kratky plot:** The Kratky plot shows that Ct_AROM shows a bell shaped curve characteristic of globular proteins. **D) Distance distribution function plot:** The $p(r)$ vs. r (in \AA) shows that the maximum dimension of the particle in solution, D_{\max} . $P(r)$ drops to zero on the r axis.

Table 3.6 Data quality and analysis parameters of Ct_AROM: The table shows the data quality analysis of SAXS data, the R_g and I_0 estimates

Total quality estimate	0.75
Guinier R_g/I_0	54.75/0.21
$P(r)$ R_g/I_0	54.74/0.21
D_{\max}	165.0 \AA
Porod volume	609000.0 \AA^3
Molecular weight (predicted/real/%error) (kDa)	318/344/7.5%

In SAXS-based modeling, SAXS profiles are computed for a given atomic structure and are compared with the experimental profile [248, 249]. The experimental SAXS curves were fitted to the computed SAXS profiles of the crystal structure of Ct_AROM (Figure 3.36). As seen in Figure 3.36 A, the calculated SAXS profile fits the experimental data very well, with a χ value of 1.92. The only region which does not fit the data completely, is the one in the q-range from 0.08 to 0.1 \AA^{-1} . Therefore, it is evident that the Ct_AROM crystal explains the conformation in solution to a very good extent.

To refine the fit, theoretical SAXS profiles were also computed for all constitutive models of the CS_AROM ensemble (Section 3.7.5), and compared to the experimental SAXS profile. We thus identified AROM conformations that explain the SAXS data better. The best fit was achieved with the model 1nva-3roi-4y0a-3fbt-D_ with a $\chi = 1.26$. Especially, the fit in the region between 0.08 to 0.1 \AA^{-1} improved significantly. In the structure of 1nva-3roi-4y0a-3fbt-D_ AROM, the AROM complex is found in a slightly more relaxed conformation, with the A domains in a more distal orientation and a larger cleft between the A and K domains. In contrast, the worst fit was achievable with the model B_ AROM-1g6t-2iyt-3don-D_ AROM (Figure 3.36 C) ($\chi = 2.12$).

Further, SAXS experiments were performed in presence of various ligands. Figure 3.36 D, E and F show the experimental SAXS profile of Ct_AROM in the presence of SKM, ATP, NADH and DHQ. This experimental profile was also fitted to computed profile from the crystal structure of Ct_AROM (Figure 3.36 D) ($\chi = 2.69$). The fit to the experimental data improved with model 1nva-A_ AROM-3mrs-3don-D_ AROM (Figure 3.36 E) ($\chi = 1.18$), which shows the A domain in a more extended conformation, while it deteriorated when fitted to the model B_ AROM-1g6t-2iyt-3don-D_ AROM (Figure 3.36 F) ($\chi = 3.14$). The crystal structures constituting these models are listed in Table 3.7.

For a surface representation, an outline of the conformational space (CS_ AROM) was constructed in Pymol. This outline is superimposed with the crystal structure and the model 1nva-2bjb-3mrs-3fbt-D_ AROM, which fit the SAXS data best until a cutoff of 0.2 \AA^{-1} . (Figure 3.37). The model occupies a greater area in the possible space of CS_ AROM and it is apparent that it shows the AROM complex in a more relaxed conformation than the crystal structure.

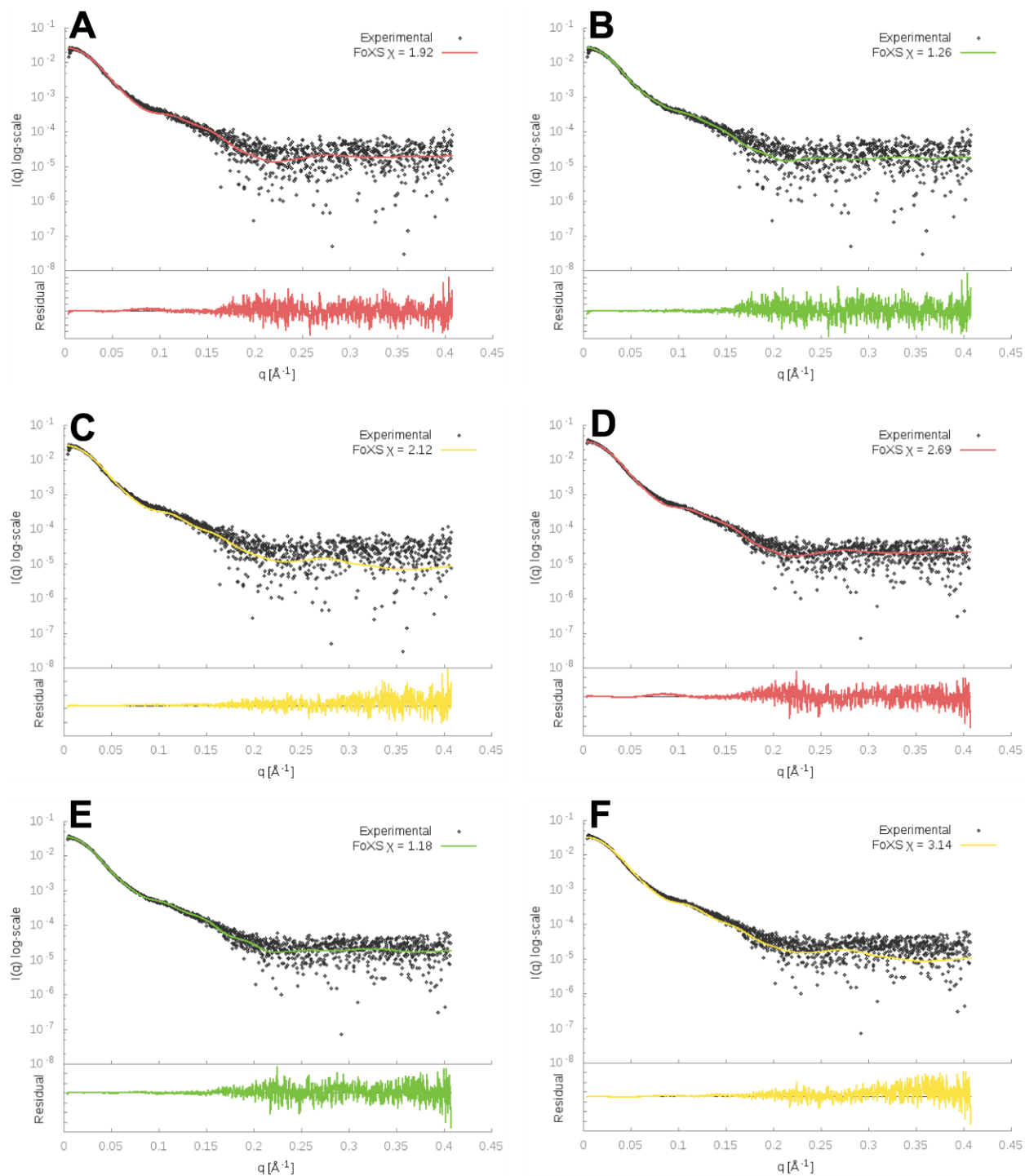


Figure 3.36 SAXS analysis: The figure shows the experimental SAXS profile (colored) superimposed on the computed SAXS profile. The curves are sorted based on the chi values. Red color shows the computed SAXS profile based on the crystal structure of Ct_AROM, Green color shows the best fitting computed SAXS profile, while yellow color shows the worst fitting computed SAXS profile. **A.) SAXS profile of Ct_AROM apo and crystal structure of Ct_AROM:** The figure compares the experimental SAXS profile of Ct_AROM apo to the computed SAXS profile based on crystal structure of Ct_AROM. **B.) SAXS profile of Ct_AROM apo and model of 1nva-3roi-4y0a-3fht-D_AROM:** The figure compares the experimental SAXS profile of Ct_AROM apo, to the computed SAXS profile from the model constituting the crystal

structures with PDBIDs: 1nva, 3roi, 4y0a, 3ft, D_AROM. **C.) SAXS profile of Ct_AROM apo and model of B_AROM-1g6t-2iyt-3don-D_AROM:** The figure compares the experimental SAXS profile of Ct_AROM apo, to the computed SAXS profile from the model constituting the crystal structures with PDBIDs: B_AROM-1g6t-2iyt-3don-D_AROM. **D.) SAXS profile of Ct_AROM +SKM+ATP+NADH+DHQ+DHS and crystal structure of Ct_AROM:** The figure compares the experimental SAXS profile of Ct_AROM+SKM+ATP+NADH+DHQ+DHS to the computed SAXS profile based on the crystal structure of Ct_AROM. **E.) SAXS profile of Ct_AROM +SKM+ATP+NADH+DHQ+DHS and model of 1nva-A_AROM-3mrs-3don-D_AROM:** The figure compares the experimental SAXS profile of Ct_AROM+SKM+ATP+NADH+DHQ+DHS to the computed SAXS profile model constituting the crystal structures with PDBIDs:1nva-A_AROM-3mrs-3don-D_AROM. **F.) SAXS profile of Ct_AROM +SKM+ATP+NADH+DHQ+DHS and model of B_AROM-1g6t-2iyt-3don-D_AROM:** The Figure compares the experimental SAXS profile of Ct_AROM+SKM+ATP+NADH+DHQ+DHS to the computed SAXS profile model constituting the crystal structures with PDBIDs: B_AROM-1g6t-2iyt-3don-D_AROM.

Table 3.7 Crystal structures employed for models

PDB ID	Domain	Ligands bound	Organism	Reference
1nva	AroB	Zn ²⁺ , ADP	<i>Aspergillus niger</i>	[91]
3roi	AroA	SO ₄ ²⁻ , Cl ⁻	<i>Coxiella burnetii</i>	[250]
3ft	AroME	-	<i>Clostridium acetobutylicum</i>	[251]
3mrs	AroK, R57A	-	<i>Helicobacter pylori</i>	[224]
1g6t	AroA	EPSP	<i>Escherichia coli</i>	[140]
2iyt	AroK	Cl ⁻	<i>Mycobacterium tuberculosis</i>	[133]
3don	AroE	Glycerol	<i>Staphylococcus epidermidis</i>	[226]
2bjb	AroA	-	<i>Mycobacterium tuberculosis</i>	[252]
4y0a	AroK	SO ₄ ²⁻ , SKM	<i>Acinetobacter baumannii</i>	[134]

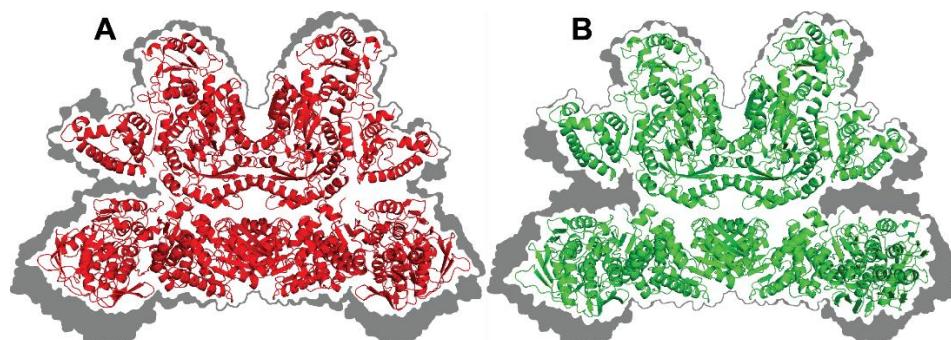


Figure 3.37. Crystal structure of Ct_AROM and model 1nva-2bjb-3mrs-3ft-DAROM superimposed on the outline of CS_AROM: The figure compares the superimposition of (A.) the crystal structure of Ct_AROM (red) (B.) and model 1nva-2bjb-3mrs-3ft-DAROM (green) in an outline of the surface representation of CS_AROM (gray).

These results suggest that Ct_AROM exists in very similar conformations in the crystal structure and in solution. While the core, containing the D domain and N-terminal B subdomain, is generally rigid the C-terminal B subdomain and the E, K and A domains can sample different conformations for different domains depending on the ligand state. This is most evident in the fit of different AROM conformations to the SAXS profiles in different ligandation states.

3.7.7 Implications of the AROM structure

The analysis of the conformational space indicated that almost all conformations that occur throughout the individual catalytic cycles can be adopted by the individual domains without causing steric clashes, and without the need for coordination between the domains. Thus, the complex appears to be optimized to concentrate all active sites in the closest possible proximity to each other. This is in accordance with the results of the kinetic analysis, which suggested an increased catalytic throughput but could not reveal regulatory interactions. This tight spatial arrangement could potentially be exploited pharmaceutically. The pentafunctional AROM complex is present in pathogenic fungi and in various apicomplexan parasites, such as *T. gondii*, and is therefore a relevant drug target [35, 37, 62, 76, 163, 203, 253, 254]. Until now, the only inhibitor used as herbicide is glyphosate [142], now, new drugs can be designed by aiming at the flexibility between the individual domains, with the aim to restrict the conformational space with an allosteric inhibitor. To this end, the defined conformational space of the AROM complex could be targeted directly in *in silico* screening. Further, since Ct_ARAM is very robust, it could pose a good model system for high throughput screening for AROM inhibitors and fragment based drug discovery.

Further, based on the crystal structure of the AROM complex, *in silico* models of related multienzymes can be generated. One of them is the tetrafunctional AroN from the pathogen *A. castalleni* (Figure 3.38). It is a 130 kDa homolog of the AROM complex that only lacks the C-terminal E domain [37]. The absence of this domain from AroN is highly remarkable and could suggest the formation of higher order assemblies. However, also in absence of this domain, the remaining four domains still provide a large conformational space for *in silico* drug screening.

The other protein that can be modeled based on the AROM structure to large extents is the QutR quinate repressor. It consists of the three C-terminal KDE domains [211] and an N-terminal part responsible for the repressor function. Canonically, this N-terminal part interacts with the quinate activator, to alter its DNA binding behavior. The QutR protein is of particular interest, as it has an intramolecular signaling function. Here, the ligation state of the KDE domains controls the association or dissociation with the quinate activator. It may therefore pose an interesting study object to follow allosteric regulatory effects mediated by the KDE domains.

Lastly, the sophisticated architecture and the apparent functional gain of the AROM complex, together with the presence of many other multifunctional enzymes in the shikimate pathway, foster

the idea that the monofunctional enzymes of the pathway also form transient complexes. These assemblies might resemble the AROM complex or they might resemble part of the models we generated. So far, the only reported such assembly is that of AroB from *B subtilis*, which supposedly forms a complex with AroC and NADPH reductase [255]. This in turn fuels the hypothesis that also AROM might form complexes with further enzymes, like AroC.

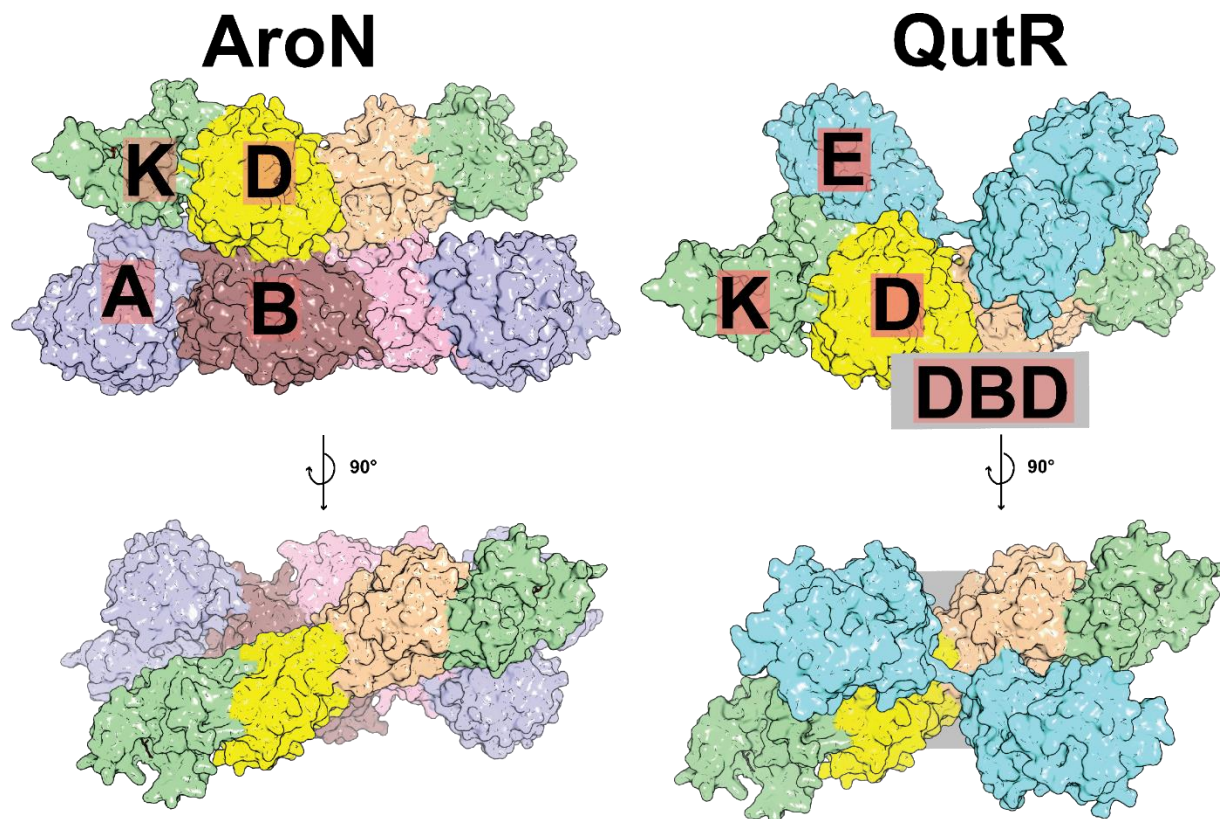


Figure 3.38. *In silico* models of AroN and QutR: The figure shows the hypothetical models of AroN, which lacks the E domain, and QutR, which contains the KDE domains, and connects N-terminally to a DNA binding domain.

4. Conclusions and Outlook

The aim of this work was to study multifunctional enzymes of the shikimate pathway, to investigate any substrate channeling or reveal regulatory mechanisms, or the formation of higher-order complexes. Much work has been done on the monofunctional enzymes of the pathway, but the knowledge on multifunctional enzymes is scarce. In this study, three bifunctional shikimate pathway enzymes (AroKB, AroEK, and AroKE) and the pentafunctional AROM complex were selected for functional and structural characterization.

The AroKB fusion catalyzes the second and fifth step of the pathway, therefore, a channeling effect cannot be anticipated, but a regulatory signaling mechanism between the domains seems conceivable. The kinetic analysis of Tm_AroKB shows that the kinetics parameters are in a similar range as reported for monofunctional bacterial AroKs and AroBs. Interestingly, the kinetic parameters of the truncated monofunctional Tm_AroK_{KB}, which is only slightly less thermostable than full-length Tm_AroKB, are also in the same range as for the bifunctional full-length protein. A significant allosteric regulation between the two enzymatic domains was not detected. Consequently, this especially wide-spread fusion of AroK and AroB domains could merely be a consequence of their genetic neighborhood within the same operon in many taxonomic groups. However, it seems possible that the fusion might be a part of higher order assemblies.

Both the AroEK and AroKE fusions catalyze subsequent steps of the pathway, therefore, a channeling effect is probable. The kinetic analysis of the K domain of Fp_AroEK yielded similar parameters as for Tm_AroKB. On the other hand, the K domain of Mb_AroKE was found to be considerably slower. Interestingly, Mb also possesses a second, monofunctional AroK. Therefore, probably this monofunctional AroK plays a dominant role in shikimate pathway, while the K domain of Mb_AroKE might perform other activities. A detailed analysis of potential channeling effects is an interesting perspective for future studies.

The crystal structures of the three bifunctional proteins yielded important and partly unexpected insights. In each of the three bifunctional proteins, the two constitutive domains are connected via rigid interfaces, such that the active sites are accessible to the solvent.

The crystal structure of Tm_AroKB is unique because it contains an intrinsically bound NAD and Zn²⁺ in the B domain. This NAD is trapped in its binding pocket by a salt bridge between Arg-

195 and Asp-316 that is specific to Tm_AroKB. A Tm_AroKB variant without the entrapped NAD was successfully designed by mutating these residues to Ala (Tm_AroKB_SB). As NAD is recycled in the AroB reaction cycle, the entrapment of this cofactor could potentially serve to increase the catalytic efficiency and potentially add to the thermostability of the protein. While the latter could be ruled out, a clear trend for the catalytic parameters could not be ascertained.

In the crystal structures of Fp_AroEK and Mb_AroKE, the individual domains had the architecture expected from the monofunctional orthologs. However, the constitutive domains in the two structures are arranged differently. Based on the two different arrangements of E and K domains found in Fp_AroEK and Mb_AroKE, two different hypothetical *in silico* models of tetrafunctional complexes were constructed by simple superposition to the Tm_AroKB and the previously reported At_AroDE structure. While both these models differ in their domain arrangement, both are physically plausible as all the catalytic sites are solvent accessible and no steric clashes were encountered. These could suggest potential transient higher order complexes that might serve to increase the overall performance of the pathway. To test for the existence of such complexes, pull down experiments in the native environment of the different proteins are a potential starting point for future studies.

In comparison to the bifunctional proteins, a functional and structural characterization of the AROM complex was performed. To circumvent stability issues reported in previous studies, the protein was taken from a thermophilic organism. Ct_AROM is a highly stable and soluble protein. In kinetic assays, its catalytic activity increased with temperature. In a simple comparison to a mixture of monofunctional enzymes from *E coli*, Ct_AROM displayed a significantly higher turnover rate. This suggests that the architecture of AROM facilitates substrate channeling. However, this must be verified by carrying out kinetic studies using a series of functional Ct_AROM mutants.

Finally, the crystal structure of Ct_AROM and its conformational space was solved and analyzed. The crystal structure initially revealed the domain arrangement, which is different from both tetrafunctional *in silico* models. All active sites are accessible to the solvent and are arranged for short diffusion times between domains responsible for consecutive steps in the pathway. The overall conformation of the crystal structure could be confirmed using SAXS and XL-MS data.

To understand the dynamics of domain motions during various catalytic states, a structural ensemble of the accessible conformational states of the AROM complex was generated by mapping the conformations of the crystal structures of the different monofunctional enzymes from the PDB in various catalytic states onto the Ct_ARAM structure. With this ensemble, the conformations that best fit the SAXS data could be identified, which closely resemble the crystal structure. Further, this suggested that individual domains could adopt essentially all conformations that occur throughout the individual catalytic cycles without steric clashes to other domains, which can be demonstrated in a movie of the conformational space. All this structural information can be used to infer the structure and conformational space of the tetrafunctional AroN from *A. castalleni* and the QutR repressor.

As the AROM complex is generally present in pathogenic fungi and apicomplexan parasites, it is a highly relevant drug target. Indeed, the catalytic activity of the Ct_ARAM complex reduced in presence of glyphosate. The robust nature of Ct_ARAM could pose a good model system for high throughput screening for AROM inhibitors and fragment based drug discovery. It is conceivable that novel inhibitors targeting the conformational flexibility needed between the domains can be explored.

Further, it seems possible that *in vivo*, the AROM complex is associated with even more enzymes of the shikimate and also downstream pathways, which are yet to be identified. Consequently, future studies should also focus on the identification of potential higher order assemblies in the native environment.

References

1. Metzler, D.E., *The Scene of Action*, in *Biochemistry*. 1977, Academic Press. p. 1-46.
2. *Principles and Techniques of Biochemistry and Molecular Biology*. 7 ed. 2010, Cambridge: Cambridge University Press.
3. Cooper, G.M., *The Central Role of Enzymes as Biological Catalysts*. , in *The Cell: A Molecular Approach*. 2nd edition. 2000., Sinauer Associates: Sunderland (MA).
4. Dixon, M. and E.C. Webb, *Enzymes*. 1964: Academic Press.
5. Manchester, K.L., *Louis Pasteur (1822–1895) — chance and the prepared mind*. Trends in Biotechnology, 1995. **13**(12): p. 511-515.
6. Voet, D. and J.G. Voet, *Biochemistry, 4th Edition*. 2010: John Wiley & Sons.
7. *Eduard Buchner - Nobel Lecture*. Nobel Lectures, Chemistry 1901-1921 1966 [cited 2017 September 19]; Available from: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1907/buchner-lecture.html.
8. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. Fifth ed. 2002: Freeman, W.H.
9. Copeland, R.A., *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. 2000: Wiley.
10. *Merriam-Webster.com*. Metabolic Pathway [Dictionary] 2017 2017 [cited 2017 August 14]; Available from: <https://www.merriam-webster.com/dictionary/pathway#medicalDictionary>.
11. *Biological Magnetic Resonance Data Bank*. 19-Jan-2017 [cited 2017 August 14]; Available from: http://www.bmrb.wisc.edu/data_library/Genes/Metabolic_Pathway_table.html.
12. Lane, N., *Life Ascending: The Ten Great Inventions of Evolution*. 2009: W.W. Norton.
13. Pace, N.R., *The universal nature of biochemistry*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(3): p. 805-808.
14. Nelson, D.L., *Lehninger principles of biochemistry*. Fourth edition ed. 2005, New York Freeman, W.H.
15. *Nature.com*. [cited 2017 August 8]; Available from: <https://www.nature.com/subjects/multienzyme-complexes>.
16. *NCBI*. Multienzyme complex [cited 2017 August 15]; Available from: <https://www.ncbi.nlm.nih.gov/mesh?Db=mesh&term=Multienzyme+Complexes>.
17. Agapakis, C.M., P.M. Boyle, and P.A. Silver, *Natural strategies for the spatial optimization of metabolism in synthetic biology*. Nat Chem Biol, 2012. **8**(6): p. 527-535.
18. MacLeod, F.A., S.R. Guiot, and J.W. Costerton, *Layered structure of bacterial aggregates produced in an upflow anaerobic sludge bed and filter reactor*. Applied and Environmental Microbiology, 1990. **56**(6): p. 1598-1607.
19. Yeates, T.O., C.S. Crowley, and S. Tanaka, *Bacterial microcompartment organelles: protein shell structure and evolution*. Annu Rev Biophys, 2010. **39**: p. 185-205.
20. Behal, R.H., et al., *Pyruvate dehydrogenase multienzyme complex. Characterization of assembly intermediates by sedimentation velocity analysis*. J Biol Chem, 1994. **269**(50): p. 31372-7.
21. Lui, T. *Chemistry Libre texts*. Pyruvate dehydrogenase complex [cited 2017 August 15]; Available from: https://chem.libretexts.org/Core/Biological_Chemistry/Metabolism/Catabolism/Pyruvate_Dehydrogenase_Complex.

22. Benson, D. *Chemistry Libre Texts*. Electron Transport chain [cited 2017 August 15]; Available from: https://chem.libretexts.org/Core/Biological_Chemistry/Metabolism/Catabolism/Electron_Transport_Chain.
23. NCBI. Multifunctional enzyme complex [cited 2017 August 15]; Available from: <https://www.ncbi.nlm.nih.gov/mesh/68064251>.
24. Yon-Kahn, J. and G. Hervé, *Multifunctional Enzymes, Multi-Enzymatic Complexes and Metabolic Channelling*, in *Molecular and Cellular Enzymology*. 2010, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 679-722.
25. Lomakin, I.B., Y. Xiong, and T.A. Steitz, *The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together*. *Cell*, 2007. **129**(2): p. 319-32.
26. Maier, T., et al., *Structure and function of eukaryotic fatty acid synthases*. *Q Rev Biophys*, 2010. **43**(3): p. 373-422.
27. Maier, T., M. Leibundgut, and N. Ban, *The crystal structure of a mammalian fatty acid synthase*. *Science*, 2008. **321**(5894): p. 1315-22.
28. Cheng, X.-Y., et al., *A Global Characterization and Identification of Multifunctional Enzymes*. *PLoS ONE*, 2012. **7**(6): p. e38979.
29. Moore, B.D., *Bifunctional and moonlighting enzymes: lighting the way to regulatory control*. *Trends in Plant Science*, 2004. **9**(5): p. 221-228.
30. Lawrence, M.C., et al., *The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of Saccharomyces cerevisiae*. *J Mol Biol*, 2005. **348**(3): p. 655-70.
31. Mousdale, D.M., M.S. Campbell, and J.R. Coggins, *Purification and characterization of bifunctional dehydroquinase-shikimate: NADP oxidoreductase from pea seedlings*. *Phytochemistry*, 1987. **26**(10): p. 2665-2670.
32. Bischoff, M., et al., *Molecular Characterization of Tomato 3-Dehydroquinase Dehydratase-Shikimate:NADP Oxidoreductase*. *Plant Physiology*, 2001. **125**(4): p. 1891-1900.
33. Singh, S.A. and D. Christendat, *Structure of Arabidopsis dehydroquinase dehydratase-shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway*. *Biochemistry*, 2006. **45**(25): p. 7787-96.
34. Ding, L., et al., *Functional analysis of the essential bifunctional tobacco enzyme 3-dehydroquinase dehydratase/shikimate dehydrogenase in transgenic tobacco plants*. *J Exp Bot*, 2007. **58**(8): p. 2053-67.
35. Richards, T.A., et al., *Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements*. *Eukaryot Cell*, 2006. **5**(9): p. 1517-31.
36. Campbell, S.A., et al., *A complete shikimate pathway in Toxoplasma gondii: an ancient eukaryotic innovation*. *Int J Parasitol*, 2004. **34**(1): p. 5-13.
37. Henriquez, F.L., et al., *The Acanthamoeba shikimate pathway has a unique molecular arrangement and is essential for aromatic amino acid biosynthesis*. *Protist*, 2015. **166**(1): p. 93-105.
38. Mir, R., S. Jallu, and T.P. Singh, *The shikimate pathway: review of amino acid sequence, function and three-dimensional structures of the enzymes*. *Crit Rev Microbiol*, 2015. **41**(2): p. 172-89.

39. Hermann, K.M. and L.M. Weaver, *The Shikimate Pathway*. Annu Rev Plant Physiol Plant Mol Biol., 1999(50): p. 473-503.
40. Bentley, R., *The shikimate pathway--a metabolic tree with many branches*. Crit Rev Biochem Mol Biol, 1990. **25**(5): p. 307-84.
41. Seigler, D.S., *Shikimic Acid Pathway*, in *Plant Secondary Metabolism*, D.S. Seigler, Editor. 1998, Springer US: Boston, MA. p. 94-105.
42. Tzin, V., G. Galili, and A. Aharoni, *Shikimate Pathway and Aromatic Amino Acid Biosynthesis*, in *eLS*. 2001, John Wiley & Sons, Ltd.
43. Maeda, H. and N. Dudareva, *The shikimate pathway and aromatic amino Acid biosynthesis in plants*. Annu Rev Plant Biol, 2012. **63**: p. 73-105.
44. Facchini, P.J., K.L. Huber-Allanach, and L.W. Tari, *Plant aromatic L-amino acid decarboxylases: evolution, biochemistry, regulation, and metabolic engineering applications*. Phytochemistry, 2000. **54**(2): p. 121-38.
45. Wakasa, K. and A. Ishihara, *Metabolic engineering of the tryptophan and phenylalanine biosynthetic pathways in rice*. Plant Biotechnology, 2009. **26**(5): p. 523-533.
46. Tzin, V. and G. Galili, *New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants*. Mol Plant, 2010. **3**(6): p. 956-72.
47. Karpf, M. and R. Trussardi, *New, azide-free transformation of epoxides into 1,2-diamino compounds: synthesis of the anti-influenza neuraminidase inhibitor oseltamivir phosphate (Tamiflu)*. J Org Chem, 2001. **66**(6): p. 2044-51.
48. Rohr, J., *Shikimic Acid. Metabolism and Metabolites*. Von E. Haslam. Wiley, Chichester, 1993. 387 S., geb. 75.00 £. – ISBN 0-471-93999-4. Angewandte Chemie, 1995. **107**(5): p. 653-653.
49. Coggins, J.R., et al., *The anatomy of a multifunctional enzyme*. Biochem Soc Trans, 1987. **15**(4): p. 754-9.
50. Buchholz, B., et al., *Reconstruction of amino acid biosynthesis by combining spinach chloroplasts with other leaf organelles*. Phytochemistry, 1979. **18**(18): p. 1109–1111.
51. Ganson, R.J., T.A. D'Amato, and R.A. Jensen, *The Two-Isozyme System of 3-Deoxy-d-arabino-Heptulosonate 7-Phosphate Synthase in Nicotiana silvestris and Other Higher Plants*. Plant Physiol, 1986. **82**(1): p. 203-10.
52. d'Amato, T.A., et al., *Subcellular localization of chorismate-mutase isoenzymes in protoplasts from mesophyll and suspension-cultured cells of Nicotiana silvestris*. Planta 1984 **162**: p. 104-108.
53. Mousdale, D.M. and J.R. Coggins, *Purification and properties of 5-enolpyruvylshikimate 3-phosphate synthase from seedlings of Pisum sativum L*. Planta, 1984. **160**(1): p. 78-83.
54. Mustafa, N.R. and R. Verpoorte, *Chorismate derived C6C1 compounds in plants*. Planta, 2005. **222**(1): p. 1-5.
55. Weber, A.P., R. Schwacke, and U.I. Flugge, *Solute transporters of the plastid envelope membrane*. Annu Rev Plant Biol, 2005. **56**: p. 133-64.
56. Della-Cioppa, G., et al., *Translocation of the precursor of 5-enolpyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro*. Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(18): p. 6873-6877.
57. Schmid, J., et al., *The in-vitro synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts*. Plant J, 1992. **2**(3): p. 375-83.

58. Galili, G. and R. Hofgen, *Metabolic engineering of amino acids and storage proteins in plants*. *Metab Eng*, 2002. **4**(1): p. 3-11.
59. Li, J. and R.L. Last, *The Arabidopsis thaliana trp5 mutant has a feedback-resistant anthranilate synthase and elevated soluble tryptophan*. *Plant Physiol*, 1996. **110**(1): p. 51-9.
60. Hawkins, A.R., et al., *The pre-chorismate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects*. *J Gen Microbiol.*, 1993. **139**(12): p. 2891-9.
61. Carpenter, E.P., et al., *Structure of dehydroquinase reveals an active site capable of multistep catalysis*. *Nature*, 1998 **394**(6690): p. 299-302.
62. Roberts, C.W., et al., *The shikimate pathway and its branches in apicomplexan parasites*. *J Infect Dis*, 2002. **185 Suppl 1**: p. S25-36.
63. Kishore, G.M. and D.M. Shah, *Amino acid biosynthesis inhibitors as herbicides*. *Annu Rev Biochem*, 1988. **57**: p. 627-63.
64. Coggins, J., et al., *Experiences with the shikimate-pathway enzymes as targets for rational drug design*. *Biochemical Society Transactions*, 2003. **31**(3): p. 548-52.
65. McConkey, G.A., *Targeting the shikimate pathway in the malaria parasite Plasmodium falciparum*. *Antimicrob Agents Chemother*, 1999. **43**(1): p. 175-7.
66. Parish, T. and N.G. Stoker, *The common aromatic amino acid biosynthesis pathway is essential in Mycobacterium tuberculosis*. *Microbiology*, 2002. **148**(10): p. 3069-77.
67. Günel-Ozcan, A., et al., *Salmonella typhimurium aroB mutants are attenuated in BALB/c mice*. *Microb Pathog.*, 1997 **23**(5): p. 311-6.
68. Segura-Cabrera, A. and M.A. Rodriguez-Perez, *Structure-based prediction of Mycobacterium tuberculosis shikimate kinase inhibitors by high-throughput virtual screening*. *Bioorg Med Chem Lett*, 2008. **18**(11): p. 3152-7.
69. Kumar, M., et al., *Structure-based in silico design of a high-affinity dipeptide inhibitor for novel protein drug target Shikimate kinase of Mycobacterium tuberculosis*. *Chem Biol Drug Des*, 2010. **76**(3): p. 277-84.
70. Saidenberg, D.M., et al., *Shikimate kinase (EC 2.7.1.71) from Mycobacterium tuberculosis: kinetics and structural dynamics of a potential molecular target for drug development*. *Curr Med Chem.*, 2011. **18**(9): p. 1299-310.
71. Reichau, S., et al., *Potent Inhibitors of a Shikimate Pathway Enzyme from Mycobacterium tuberculosis: COMBINING MECHANISM- AND MODELING-BASED DESIGN*. *The Journal of Biological Chemistry*, 2011. **286**(18): p. 16197-16207.
72. Blanco, B., et al., *Mycobacterium tuberculosis shikimate kinase inhibitors: design and simulation studies of the catalytic turnover*. *J Am Chem Soc*, 2013. **135**(33): p. 12366-76.
73. Lamichhane, G., et al., *Essential Metabolites of Mycobacterium tuberculosis and Their Mimics*. *mBio*, 2011. **2**(1): p. e00301-10.
74. Chim, N., et al., *Advances in Mycobacterium tuberculosis therapeutics discovery utilizing structural biology*. *Infectious disorders drug targets*, 2012.
75. Eschenburg, S., et al., *A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states*. *J Biol Chem*, 2003. **278**(49): p. 49215-22.
76. Roberts, F., et al., *Evidence for the shikimate pathway in apicomplexan parasites*. *Nature*, 1998. **393**(6687): p. 801-805.

77. Welch, G.R. and F.H. Gaertner, *Coordinate activation of a multienzyme complex by the first substrate. Evidence for a novel regulatory mechanism in the polyaromatic pathway of Neurospora crassa.* Arch Biochem Biophys, 1976. **172**(2): p. 476-89.
78. Lumsden, J. and J.R. Coggins, *The subunit structure of the arom multienzyme complex of Neurospora crassa. Evidence from peptide 'maps' for the identity of the subunits.* Biochem J, 1978. **169**(2): p. 441-4.
79. Patel, V.B. and N.H. Giles, *Purification of the arom multienzyme aggregate from Euglena gracilis.* Biochim Biophys Acta, 1979. **567**(1): p. 24-34.
80. Garner, C.C. and K.M. Herrmann, *Structural analysis of 3-deoxy-D-arabino-heptulosonate 7-phosphate by 1H- and natural-abundance 13C-n.m.r. spectroscopy.* Carbohydr Res, 1984. **132**(2): p. 317-22.
81. Huisman, O.C. and T. Kosuge, *Regulation of aromatic amino acid biosynthesis in higher plants. II. 3-Deoxy-arabino-heptulosonic acid 7-phosphate synthetase from cauliflower.* J Biol Chem, 1974. **249**(21): p. 6842-8.
82. Rubin, J.L. and R.A. Jensen, *Differentially Regulated Isozymes of 3-Deoxy-d-arabino-Heptulosonate-7-Phosphate Synthase from Seedlings of Vigna radiata [L.] Wilczek.* Plant Physiol, 1985. **79**(3): p. 711-8.
83. Mukhopadhyay, S., M.S. Hasson, and D.A. Sanders, *A continuous assay of acetate kinase activity: measurement of inorganic phosphate release generated by hydroxylaminolysis of acetyl phosphate.* Bioorg Chem, 2008. **36**(2): p. 65-9.
84. Webby, C.J., et al., *The structure of 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase from Mycobacterium tuberculosis reveals a common catalytic scaffold and ancestry for type I and type II enzymes.* J Mol Biol, 2005. **354**(4): p. 927-39.
85. Dyer, W.E., et al., *A cDNA encoding 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from Solanum tuberosum L.* Journal of Biological Chemistry, 1990. **265**(3): p. 1608-14.
86. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**: p. 537.
87. Baasov, T. and J.R. Knowles, *Is the first enzyme of the shikimate pathway, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (tyrosine sensitive), a copper metalloenzyme?* Journal of Bacteriology, 1989. **171**(11): p. 6155-6160.
88. Jensen, R.A. and D.S. Nasser, *Comparative Regulation of Isoenzymic 3-Deoxy-d-arabino-Heptulosonate 7-Phosphate Synthetases in Microorganisms.* Journal of Bacteriology, 1968. **95**(1): p. 188-196.
89. de Mendonça, J.D., et al., *Functional Characterization by Genetic Complementation of aroB-Encoded Dehydroquinase Synthase from Mycobacterium tuberculosis H37Rv and Its Heterologous Expression and Purification.* Journal of Bacteriology, 2007. **189**(17): p. 6246-6252.
90. Floss, H.G., D.K. Onderka, and M. Carroll, *Stereochemistry of the 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase reaction and the chorismate synthetase reaction.* J Biol Chem, 1972. **247**(3): p. 736-44.
91. Nichols, C.E., et al., *Ligand-induced conformational changes and a mechanism for domain closure in Aspergillus nidulans dehydroquinase synthase.* J Mol Biol, 2003. **327**(1): p. 129-44.

92. Nichols, C.E., et al., *Comparison of ligand-induced conformational changes and domain closure mechanisms, between prokaryotic and eukaryotic dehydroquinase synthases*. J Mol Biol, 2004. **343**(3): p. 533-46.
93. Haslam, E., *The Shikimate Pathway*. John Wiley & Sons, Inc., 1974.
94. Chaudhuri, S., et al., *Purification and characterization of 3-dehydroquinase from Escherichia coli*. Biochemical Journal, 1986. **239**(3): p. 699-704.
95. Moore, J.D., et al., *Characterization of the type I dehydroquinase from Salmonella typhi*. Biochemical Journal, 1993. **295**(Pt 1): p. 277-285.
96. Moore, J.D., et al., *Inducible overproduction of the Aspergillus nidulans pentafunctional AROM protein and the type-I and -II 3-dehydroquinases from Salmonella typhi and Mycobacterium tuberculosis*. Biochemical Journal, 1992. **287**(Pt 1): p. 173-181.
97. Garbe, T., et al., *The Mycobacterium tuberculosis shikimate pathway genes: evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases*. Mol Gen Genet, 1991. **228**(3): p. 385-92.
98. White, P.J., et al., *The purification and characterization of 3-dehydroquinase from Streptomyces coelicolor*. Biochem J, 1990. **265**(3): p. 735-8.
99. Bottomley, J.R., et al., *Cloning, sequencing, expression, purification and preliminary characterization of a type II dehydroquinase from Helicobacter pylori*. Biochem J, 1996. **319** (Pt 2): p. 559-65.
100. Giles, N.H., et al., *Gene organization and regulation in the qa (quinic acid) gene cluster of Neurospora crassa*. Microbiological Reviews, 1985. **49**(3): p. 338-358.
101. Grant, S., et al., *Genetic regulation of the quinic acid utilization (QUT) gene cluster in Aspergillus nidulans*. J Gen Microbiol, 1988. **134**(2): p. 347-58.
102. Euverink, G.J.W., et al., *Purification and characterization of a dual function 3-dehydroquinase dehydratase from Amycolatopsis methanolica* Journal of General Microbiology, 1992. **138**: p. 2449-2457.
103. Kleanthous, C., et al., *A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes*. Biochemical Journal, 1992. **282**(3): p. 687-695.
104. Shneier, A., et al., *Observation of an imine intermediate on dehydroquinase by electrospray mass spectrometry*. Journal of the American Chemical Society, 1991. **113**(24): p. 9416-9418.
105. Shneier, A., et al., *Evidence for opposite stereochemical courses for the reaction catalysed by type I and type II dehydroquinases*. Bioorganic & Medicinal Chemistry Letters, 1993. **3**(7): p. 1399-1402.
106. Butler, J.R., W.L. Alworth, and M.J. Nugent, *Mechanism of dehydroquinase catalyzed dehydration. I. Formation of a Schiff base intermediate*. Journal of the American Chemical Society, 1974. **96**(5): p. 1617-1618.
107. Walsh, C., *Enzymatic Reaction Mechanisms*. 1979: W. H. Freeman.
108. Yao, Y. and Z.S. Li, *New insights into the mechanism of the Schiff base hydrolysis catalyzed by type I dehydroquinase dehydratase from S. enterica: a theoretical study*. Org Biomol Chem, 2012. **10**(35): p. 7037-44.
109. Light, S.H., W.F. Anderson, and A. Lavie, *Reassessing the type I dehydroquinase dehydratase catalytic triad: kinetic and structural studies of Glu86 mutants*. Protein Sci, 2013. **22**(4): p. 418-24.

110. Harris, J., et al., *Different mechanistic and stereochemical courses for the reactions catalysed by type I and type II dehydroquinases*. Journal of the Chemical Society, Chemical Communications, 1993(13): p. 1080-1081.
111. Harris, J.M., et al., *Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases*. Biochemical Journal, 1996. **319**(Pt 2): p. 333-336.
112. Nichols, C.E., et al., *Crystal structures of Staphylococcus aureus type I dehydroquinase from enzyme turnover experiments*. Proteins, 2004. **56**(3): p. 625-8.
113. Gourley, D.G., et al., *The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction*. Nat Struct Biol, 1999. **6**(6): p. 521-5.
114. Light, S.H., et al., *Crystal Structures of Type I Dehydroquinase Dehydratase in Complex with Quinate and Shikimate Suggest a Novel Mechanism of Schiff Base Formation*. Biochemistry, 2014. **53**(5): p. 872-880.
115. Peek, J., et al., *Structural and mechanistic analysis of a novel class of shikimate dehydrogenases: evidence for a conserved catalytic mechanism in the shikimate dehydrogenase family*. Biochemistry, 2011. **50**(40): p. 8616-27.
116. Michel, G., et al., *Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities*. J Biol Chem, 2003. **278**(21): p. 19463-72.
117. Singh, S., et al., *Crystal structure of a novel shikimate dehydrogenase from Haemophilus influenzae*. J Biol Chem, 2005. **280**(17): p. 17101-8.
118. Singh, S., et al., *A phylogenomic analysis of the shikimate dehydrogenases reveals broadscale functional diversification and identifies one functionally distinct subclass*. Mol Biol Evol, 2008. **25**(10): p. 2221-32.
119. Ye, S., et al., *The Crystal Structure of Shikimate Dehydrogenase (AroE) Reveals a Unique NADPH Binding Mode*. Journal of Bacteriology, 2003. **185**(14): p. 4144-4151.
120. Lesk, A.M., *NAD-binding domains of dehydrogenases*. Curr Opin Struct Biol, 1995. **5**(6): p. 775-83.
121. Huang, L., A.L. Montoya, and E.W. Nester, *Purification and characterization of shikimate kinase enzyme activity in Bacillus subtilis*. J Biol Chem, 1975. **250**(19): p. 7675-81.
122. Lobner-Olesen, A. and M.G. Marinus, *Identification of the gene (aroK) encoding shikimate acid kinase I of Escherichia coli*. J Bacteriol, 1992. **174**(2): p. 525-9.
123. Griffin, H.G. and M.J. Gasson, *The gene (aroK) encoding shikimate kinase I from Escherichia coli*. DNA Seq, 1995. **5**(3): p. 195-7.
124. Whipp, M.J. and A.J. Pittard, *A reassessment of the relationship between aroK- and aroL-encoded shikimate kinase enzymes of Escherichia coli*. J Bacteriol, 1995. **177**(6): p. 1627-9.
125. De Feyter, R., *Shikimate kinases from Escherichia coli K12*. Methods Enzymol, 1987. **142**: p. 355-61.
126. Fucile, G., S. Falconer, and D. Christendat, *Evolutionary diversification of plant shikimate kinase gene duplicates*. PLoS Genet, 2008. **4**(12): p. e1000292.
127. Fucile, G., et al., *Structural and biochemical investigation of two Arabidopsis shikimate kinases: the heat-inducible isoform is thermostable*. Protein Sci, 2011. **20**(7): p. 1125-36.
128. Kasai, K., et al., *Identification of three shikimate kinase genes in rice: characterization of their differential expression during panicle development and of the enzymatic activities of the encoded proteins*. Planta, 2005. **222**(3): p. 438-47.

129. Krell, T., J.R. Coggins, and A.J. Laphorn, *The three-dimensional structure of shikimate kinase*. J Mol Biol, 1998. **278**(5): p. 983-97.
130. Romanowski, M.J. and S.K. Burley, *Crystal structure of the Escherichia coli shikimate kinase I (AroK) that confers sensitivity to mecillinam*. Proteins, 2002. **47**(4): p. 558-62.
131. Gu, Y., et al., *Crystal structure of shikimate kinase from Mycobacterium tuberculosis reveals the dynamic role of the LID domain in catalysis*. J Mol Biol, 2002. **319**(3): p. 779-89.
132. Gan, J., et al., *Crystal structure of Mycobacterium tuberculosis shikimate kinase in complex with shikimic acid and an ATP analogue*. Biochemistry, 2006. **45**(28): p. 8539-45.
133. Hartmann, M.D., et al., *Mechanism of phosphoryl transfer catalyzed by shikimate kinase from Mycobacterium tuberculosis*. J Mol Biol, 2006. **364**(3): p. 411-23.
134. Sutton, K.A., et al., *Structure of shikimate kinase, an in vivo essential metabolic enzyme in the nosocomial pathogen Acinetobacter baumannii, in complex with shikimate*. Acta Crystallogr D Biol Crystallogr, 2015. **71**(Pt 8): p. 1736-44.
135. Leipe, D.D., E.V. Koonin, and L. Aravind, *Evolution and Classification of P-loop Kinases and Related Proteins*. Journal of Molecular Biology, 2003. **333**(4): p. 781-815.
136. Grimshaw, C.E., S.G. Sogo, and J.R. Knowles, *The fate of the hydrogens of phosphoenolpyruvate in the reaction catalyzed by 5-enolpyruvylshikimate-3-phosphate synthase. Isotope effects and isotope exchange*. Journal of Biological Chemistry, 1982. **257**(2): p. 596-598.
137. Anton, D.L., et al., *Mechanism of enolpyruvylshikimate-3-phosphate synthase exchange of phosphoenolpyruvate with solvent protons*. Biochemistry, 1983. **22**(25): p. 5903-5908.
138. Jakeman, D.L., et al., *On the Mechanism of 5-Enolpyruvylshikimate-3-phosphate Synthase*. Biochemistry, 1998. **37**(35): p. 12012-12019.
139. Radaev, S., et al., *Structure and Mechanism of 3-Deoxy-d-manno-octulosonate 8-Phosphate Synthase*. Journal of Biological Chemistry, 2000. **275**(13): p. 9476-9484.
140. Schonbrunn, E., et al., *Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail*. Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1376-80.
141. Funke, T., et al., *Molecular basis for the herbicide resistance of Roundup Ready crops*. Proc Natl Acad Sci U S A, 2006. **103**(35): p. 13010-5.
142. Duke, S.O., *The history and current status of glyphosate*. Pest Manag Sci, 2017.
143. Light, S.H., et al., *1.70 Angstrom resolution structure of 3-phosphoshikimate 1-carboxyvinyltransferase(AroA) from Coxiella burnetii in complex with shikimate-3-phosphate and glyphosate*. To be published, <http://www.rcsb.org/pdb/explore/explore.do?structureId=3SLH>.
144. Stallings, W.C., et al., *Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: a distinctive protein fold*. Proc Natl Acad Sci U S A, 1991. **88**(11): p. 5046-50.
145. Bornemann, S., et al., *Escherichia coli chorismate synthase catalyzes the conversion of (6S)-6-fluoro-5-enolpyruvylshikimate-3-phosphate to 6-fluorochorismate. Implications for the enzyme mechanism and the antimicrobial action of (6S)-6-fluoroshikimate*. J Biol Chem, 1995. **270**(39): p. 22811-5.
146. Hawkes, T.R., et al., *Chorismate synthase. Pre-steady-state kinetics of phosphate release from 5-enolpyruvylshikimate 3-phosphate*. Biochem J, 1990. **265**(3): p. 899-902.

147. Bornemann, S., D.J. Lowe, and R.N. Thorneley, *The transient kinetics of Escherichia coli chorismate synthase: substrate consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate*. *Biochemistry*, 1996. **35**(30): p. 9907-16.
148. Macheroux, P., et al., *Studies with flavin analogs provide evidence that a protonated reduced FMN is the substrate-induced transient intermediate in the reaction of Escherichia coli chorismate synthase*. *J Biol Chem*, 1996. **271**(42): p. 25850-8.
149. Hill, R.K. and G.R. Newkome, *Stereochemistry of chorismic acid biosynthesis*. *J Am Chem Soc*, 1969. **91**(21): p. 5893-4.
150. Onderka, D.K. and H.G. Floss, *Steric course of the chorismate synthetase reaction and the 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase reaction*. *J Am Chem Soc*, 1969. **91**(21): p. 5894-6.
151. Macheroux, P., et al., *A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway*. *Planta*, 1999. **207**(3): p. 325-34.
152. White, P.J., G. Millar, and J.R. Coggins, *The overexpression, purification and complete amino acid sequence of chorismate synthase from Escherichia coli K12 and its comparison with the enzyme from Neurospora crassa*. *Biochem J*, 1988. **251**(2): p. 313-22.
153. Hasan, N. and E.W. Nester, *Purification and properties of chorismate synthase from Bacillus subtilis*. *J Biol Chem*, 1978. **253**(14): p. 4993-8.
154. Hasan, N. and E.W. Nester, *Purification and characterization of NADPH-dependent flavin reductase. An enzyme required for the activation of chorismate synthase in Bacillus subtilis*. *J Biol Chem*, 1978. **253**(14): p. 4987-92.
155. Guo, J., et al., *Molecular characterization of quinate and shikimate metabolism in Populus trichocarpa*. *J Biol Chem*, 2014. **289**(34): p. 23846-58.
156. Fiedler, E. and G. Schultz, *Localization, Purification, and Characterization of Shikimate Oxidoreductase-Dehydroquinase Hydrolyase from Stroma of Spinach Chloroplasts*. *Plant Physiology*, 1985. **79**(1): p. 212-218.
157. Tatum, E.L., et al., *Synthesis of Aromatic Compounds by Neurospora*. *Proceedings of the National Academy of Sciences of the United States of America*, 1954. **40**(5): p. 271-276.
158. Giles, N.H., et al., *A gene cluster in Neurospora crassa coding for an aggregate of five aromatic synthetic enzymes*. *Proc Natl Acad Sci U S A*, 1967. **58**(4): p. 1453-60.
159. Gaertner, F.H., *Purification of two multienzyme complexes in the aromatic/tryptophan pathway of Neurospora crassa*. *Archives of Biochemistry and Biophysics*, 1972. **151**(1): p. 277-284.
160. Jacobson, J.W., et al., *Purification and stability of the multienzyme complex encoded in the arom gene cluster of Neurospora crassa*. *Biochim Biophys Acta*, 1972. **289**(1): p. 1-12.
161. Gaertner, F.H. and K.W. Cole, *A cluster-gene: evidence for one gene, one polypeptide, five enzymes*. *Biochem Biophys Res Commun*, 1977. **75**(2): p. 259-64.
162. Smith, D.D. and J.R. Coggins, *Isolation of a bifunctional domain from the pentafunctional arom enzyme complex of Neurospora crassa*. *Biochem J*, 1983. **213**(2): p. 405-15.
163. Charles, I.G., et al., *The isolation and nucleotide sequence of the complex AROM locus of Aspergillus nidulans*. *Nucleic Acids Res*, 1986. **14**(5): p. 2201-13.
164. Morrissette, N.S. and L.D. Sibley, *Cytoskeleton of Apicomplexan Parasites*. *Microbiology and Molecular Biology Reviews*, 2002. **66**(1): p. 21-38.
165. Peek, J., et al., *Isolation and molecular characterization of the shikimate dehydrogenase domain from the Toxoplasma gondii AROM complex*. *Molecular and Biochemical Parasitology*, 2014. **194**(1): p. 16-19.

166. Gaertner, F.H., M.C. Ericson, and J.A. DeMoss, *Catalytic facilitation in vitro by two multienzyme complexes from Neurospora crassa*. J Biol Chem, 1970. **245**(3): p. 595-600.
167. Duncan, K., R.M. Edwards, and J.R. Coggins, *The Saccharomyces cerevisiae ARO1 gene. An example of the co-ordinate regulation of five enzymes on a single biosynthetic pathway*. FEBS Lett, 1988. **241**(1-2): p. 83-8.
168. Staff, H.c. *History.com*. Rosetta Stone found 2009 [cited 2017 September 1]; Available from: <http://www.history.com/this-day-in-history/rosetta-stone-found>.
169. *Rosetta Stone*. [cited 2017 September 7]; Available from: <https://discoveringegypt.com/egyptian-video-documentaries/mystery-of-the-rosetta-stone/>.
170. Marcotte, E.M., et al., *Detecting protein function and protein-protein interactions from genome sequences*. Science, 1999. **285**(5428): p. 751-3.
171. Enright, A.J., et al., *Protein interaction maps for complete genomes based on gene fusion events*. Nature, 1999. **402**(6757): p. 86-90.
172. Eisenberg, D., et al., *Protein function in the post-genomic era*. Nature, 2000. **405**(6788): p. 823-6.
173. Veitia, R.A., *Rosetta Stone proteins: "chance and necessity"?* Genome Biol, 2002. **3**(2): p. Interactions1001.
174. DeLano, W.L. *The PyMOL Molecular Graphics System*. 2002; Available from: <http://www.pymol.org>.
175. Johnson, L.S., S.R. Eddy, and E. Portugaly, *Hidden Markov model speed heuristic and iterative HMM search procedure*. BMC Bioinformatics, 2010. **11**: p. 431-431.
176. Frickey, T. and A. Lupas, *CLANS: a Java application for visualizing protein families based on pairwise similarity*. Bioinformatics, 2004. **20**(18): p. 3702-4.
177. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
178. Kabsch, W., *XDS*. Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.
179. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement*. Acta Crystallographica Section D, 2010. **66**(2): p. 133-144.
180. Vagin, A. and A. Teplyakov, *MOLREP: an Automated Program for Molecular Replacement*. Journal of Applied Crystallography, 1997. **30**(6): p. 1022-1025.
181. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of Macromolecular Structures by the Maximum-Likelihood Method*. Acta Crystallographica Section D, 1997. **53**(3): p. 240-255.
182. Murshudov, G.N., et al., *Efficient anisotropic refinement of macromolecular structures using FFT*. Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 1): p. 247-55.
183. Emsley, P., et al., *Features and development of Coot*. Acta Crystallographica Section D, 2010. **66**(4): p. 486-501.
184. Hemsley, A., et al., *A simple method for site-directed mutagenesis using the polymerase chain reaction*. Nucleic Acids Res, 1989. **17**(16): p. 6545-51.
185. Kornberg, A. and W.E. Pricer, Jr., *Enzymatic phosphorylation of adenosine and 2,6-diaminopurine riboside*. J Biol Chem, 1951. **193**(2): p. 481-95.
186. Daugherty, M., et al., *Archaeal shikimate kinase, a new member of the GHMP-kinase family*. J Bacteriol, 2001. **183**(1): p. 292-300.

187. Rosado, L.A., et al., *The mode of action of recombinant Mycobacterium tuberculosis shikimate kinase: kinetics and thermodynamics analyses*. PLoS One, 2013. **8**(5): p. e61918.
188. *EnzChek® Phosphate Assay Kit*. Molecular Probes. p. 4.
189. Segel, I.H., *Biochemical calculations: how to solve mathematical problems in general biochemistry*. 1976: Wiley.
190. Segel, I.H., *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. 1993: Wiley.
191. Kirsch, J.F., *Enzyme kinetics and mechanism*, by Paul F. Cook and W.W. Cleland. Protein Science : A Publication of the Protein Society, 2008. **17**(2): p. 380-381.
192. Berg JM, T.J., Stryer L. , *Section 8.4, The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes*, in *Biochemistry*. 2002, W H Freeman: New York.
193. *Protocols Online*. M9 medium July 17, 2012 [cited 2016 April 29]; Available from: <http://www.protocolsonline.com/recipes/media/m9-medium-5x/#section-3>.
194. Fonseca, I.O., et al., *Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: Purification and characterization*. Protein Expression and Purification, 2006. **46**(2): p. 429-437.
195. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis*. Journal of Applied Crystallography, 2003. **36**(5): p. 1277-1282.
196. Forster, S., L. Apostol, and W. Bras, *Scatter: software for the analysis of nano- and mesoscale small-angle scattering*. Journal of Applied Crystallography, 2010. **43**(3): p. 639-646.
197. Schneidman-Duhovny, D., et al., *FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles*. Nucleic Acids Research, 2016. **44**(Web Server issue): p. W424-W429.
198. Sheldrick, G., *A short history of SHELX*. Acta Crystallographica Section A, 2008. **64**(1): p. 112-122.
199. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 9): p. 1002-11.
200. Nichols, C.E., A.R. Hawkins, and D.K. Stammers, *Structure of the 'open' form of Aspergillus nidulans 3-dehydroquinase synthase at 1.7 Å resolution from crystals grown following enzyme turnover*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 971-3.
201. *1.02 Ångstrom Resolution Crystal Structure of 3-Phosphoshikimate 1-Carboxyvinyltransferase from Vibrio cholerae in complex with Shikimate-3-Phosphate (Partially Photolyzed) and Glyphosate*. <http://www.rcsb.org/structure/3NVS>.
202. Hoppner, A., D. Schomburg, and K. Niefind, *Enzyme-substrate complexes of the quinase/shikimate dehydrogenase from Corynebacterium glutamicum enable new insights in substrate and cofactor binding, specificity, and discrimination*. Biol Chem, 2013. **394**(11): p. 1505-16.
203. Welch, G.R. and F.H. Gaertner, *Enzyme organization in the polyaromatic-biosynthetic pathway: the arom conjugate and other multienzyme systems*. Curr Top Cell Regul, 1980. **16**: p. 113-62.
204. Coggins, J.R., et al., *The arom multifunctional enzyme from Neurospora crassa*. Methods in Enzymology, 1987. **142**: p. 325-341.
205. Hawkins, A.R., *The complex Arom locus of Aspergillus nidulans. Evidence for multiple gene fusions and convergent evolution*. Curr Genet, 1987. **11**(6-7): p. 491-8.

206. Lamb, H.K., C.R. Bagshaw, and A.R. Hawkins, *In vivo overproduction of the pentafunctional arom polypeptide in Aspergillus nidulans affects metabolic flux in the quinate pathway*. Mol Gen Genet, 1991. **227**(2): p. 187-96.
207. Hawkins, A.R., J.D. Moore, and H.K. Lamb, *The molecular biology of the pentafunctional AROM protein*. Biochem Soc Trans, 1993. **21**(1): p. 181-6.
208. Nitsche, F., et al., *Higher level taxonomy and molecular phylogenetics of the Choanoflagellata*. J Eukaryot Microbiol, 2011. **58**(5): p. 452-62.
209. Waggoner, B. *University of California Museum of Paleontology*. Introduction to the Oomycota 12/21/95 [cited 2017 October, 10]; Available from: <http://www.ucmp.berkeley.edu/chromista/oomycota.html>.
210. Kimball, J.W. *Kimball's Biology Pages*. Protists 2017 2017 September 19 [cited 2017 October 17]; Available from: <http://www.biology-pages.info/P/Protists.html>.
211. Lamb, H.K., et al., *The QUTA activator and QUTR repressor proteins of Aspergillus nidulans interact to regulate transcription of the quinate utilization pathway genes*. Microbiology, 1996. **142** (Pt 6): p. 1477-90.
212. Anton, I.A., K. Duncan, and J.R. Coggins, *A eukaryotic repressor protein, the qa-1S gene product of Neurospora crassa, is homologous to part of the arom multifunctional enzyme*. J Mol Biol, 1987. **197**(2): p. 367-71.
213. Huber, R., et al., *Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C*. Archives of Microbiology, 1986. **144**(4): p. 324-333.
214. Kröning, N., et al., *ATP Binding to the KTN/RCK Subunit KtrA from the K⁺-uptake System KtrAB of Vibrio alginolyticus: ITS ROLE IN THE FORMATION OF THE KtrAB COMPLEX AND ITS REQUIREMENT IN VIVO*. Journal of Biological Chemistry, 2007. **282**(19): p. 14018-14027.
215. Negron, L. and E.J. Parker, *Fluorinated substrates result in variable leakage of a reaction intermediate during catalysis by dehydroquinase synthase*. Org Biomol Chem, 2011. **9**(8): p. 2861-7.
216. Park, A., et al., *Biophysical and kinetic analysis of wild-type and site-directed mutants of the isolated and native dehydroquinase synthase domain of the AROM protein*. Protein Science, 2004. **13**(8): p. 2108-2119.
217. Mehdi, S., J.W. Frost, and J.R. Knowles, *Dehydroquinase synthase from escherichia coli, and its substrate 3-deoxy-d-arabino-heptulosonic acid 7-phosphate*. Methods in Enzymology, 1987. **142**: p. 306-314.
218. Bender, S.L., S. Mehdi, and J.R. Knowles, *Dehydroquinase synthase: the role of divalent metal cations and of nicotinamide adenine dinucleotide in catalysis*. Biochemistry, 1989. **28**(19): p. 7555-60.
219. Moore, J.D., et al., *Efficient independent activity of a monomeric, monofunctional dehydroquinase synthase derived from the N-terminus of the pentafunctional AROM protein of Aspergillus nidulans*. Biochemical Journal, 1994. **301**(1): p. 297-304.
220. Negron, L., M.L. Patchett, and E.J. Parker, *Expression, Purification, and Characterisation of Dehydroquinase Synthase from Pyrococcus furiosus*. Enzyme Res, 2011. **2011**: p. 134893.
221. Mittelstadt, G., et al., *Biochemical and structural characterisation of dehydroquinase synthase from the New Zealand kiwifruit Actinidia chinensis*. Arch Biochem Biophys, 2013. **537**(2): p. 185-91.

222. Chan, C.-H., T.-H. Yu, and K.-B. Wong, *Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding*. PLOS ONE, 2011. **6**(6): p. e21624.
223. Millar, G., et al., *The cloning and expression of the aroL gene from Escherichia coli K12*. Biochem. J., 1986. **237**: p. 427-437.
224. Cheng, W.C., et al., *Structures of Helicobacter pylori shikimate kinase reveal a selective inhibitor-induced-fit mechanism*. PLoS One, 2012. **7**(3): p. e33481.
225. Miquel, S., et al., *Faecalibacterium prausnitzii and human intestinal health*. Current Opinion in Microbiology, 2013. **16**(3): p. 255-261.
226. Han, C., et al., *X-ray crystallographic and enzymatic analyses of shikimate dehydrogenase from Staphylococcus epidermidis*. Febs j, 2009. **276**(4): p. 1125-39.
227. Bagautdinov, B. and N. Kunishima, *Crystal structures of shikimate dehydrogenase AroE from Thermus thermophilus HB8 and its cofactor and substrate complexes: insights into the enzymatic mechanism*. J Mol Biol, 2007. **373**(2): p. 424-38.
228. Gan, J., et al., *Structural and biochemical analyses of shikimate dehydrogenase AroE from Aquifex aeolicus: implications for the catalytic mechanism*. Biochemistry, 2007. **46**(33): p. 9513-22.
229. Reddy Chichili, V.P., V. Kumar, and J. Sivaraman, *Linkers in the structural biology of protein-protein interactions*. Protein Science : A Publication of the Protein Society, 2013. **22**(2): p. 153-167.
230. Brauer, S.L., et al., *Methanoregula boonei gen. nov., sp. nov., an acidiphilic methanogen isolated from an acidic peat bog*. Int J Syst Evol Microbiol, 2011. **61**(Pt 1): p. 45-52.
231. Harel, A. and Y. Gruenbaum, *Nuclear Pore Structure: Warming up the Core*. Cell, 2011. **146**(2): p. 191-193.
232. Bock, T., et al., *An integrated approach for genome annotation of the eukaryotic thermophile Chaetomium thermophilum*. Nucleic Acids Research, 2014. **42**(22): p. 13525-13533.
233. van den Brink, J., et al., *Phylogeny of the industrial relevant, thermophilic genera Myceliophthora and Corynascus*. Fungal Diversity, 2012. **52**(1): p. 197-207.
234. Maheshwari, R., G. Bharadwaj, and M.K. Bhat, *Thermophilic Fungi: Their Physiology and Enzymes*. Microbiology and Molecular Biology Reviews, 2000. **64**(3): p. 461-488.
235. Lambert, J.M., M.R. Boocock, and J.R. Coggins, *The 3-dehydroquinase activity of the pentafunctional arom enzyme complex of Neurospora crassa is Zn²⁺-dependent*. Biochem J, 1985. **226**(3): p. 817-29.
236. Moore, J.D. and A.R. Hawkins, *Overproduction of, and interaction within, bifunctional domains from the amino- and carboxy-termini of the pentafunctional AROM protein of Aspergillus nidulans*. Mol Gen Genet, 1993. **240**(1): p. 92-102.
237. Boocock, M.R. and J.R. Coggins, *Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate*. FEBS Lett, 1983. **154**(1): p. 127-33.
238. Gaertner, F.H. and J.A. DeMoss, *[48] Erythrose 4-phosphate and phosphoenolpyruvate to indole-3-glycerol phosphate: Two multienzyme complexes and three separable enzymes in Neurospora crassa*, in *Methods in Enzymology*. 1970, Academic Press. p. 387-401.
239. Priestman, M.A., et al., *5-Enolpyruvylshikimate-3-phosphate synthase from Staphylococcus aureus is insensitive to glyphosate*. FEBS Letters, 2005. **579**(3): p. 728-732.

240. Chaudhary, S. and J.R. Coggins, *The purification of shikimate dehydrogenase from Escherichia coli*. *Biochem J.*, 1985. **226**(1): p. 217–223.
241. Leitner, A., et al., *Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines*. *Trends Biochem Sci*, 2016. **41**(1): p. 20-32.
242. Leitner, A., et al., *Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics*. *Molecular & Cellular Proteomics : MCP*, 2010. **9**(8): p. 1634-1649.
243. Merkley, E.D., et al., *Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances*. *Protein Science : A Publication of the Protein Society*, 2014. **23**(6): p. 747-759.
244. Priestman, M.A., et al., *Molecular basis for the glyphosate-insensitivity of the reaction of 5-enolpyruvylshikimate 3-phosphate synthase with shikimate*. *FEBS Lett*, 2005. **579**(25): p. 5773-80.
245. Funke, T., et al., *Structural basis of glyphosate resistance resulting from the double mutation Thr97 -> Ile and Pro101 -> Ser in 5-enolpyruvylshikimate-3-phosphate synthase from Escherichia coli*. *J Biol Chem*, 2009. **284**(15): p. 9854-60.
246. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins*. *FEBS Letters*, 2015. **589**(19, Part A): p. 2570-2577.
247. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution*. *Q Rev Biophys*, 2007. **40**(3): p. 191-285.
248. Schneidman-Duhovny, D., et al., *Accurate SAXS profile computation and its assessment by contrast variation experiments*. *Biophys J*, 2013. **105**(4): p. 962-74.
249. Schneidman-Duhovny, D., M. Hammel, and A. Sali, *FoXS: a web server for rapid computation and fitting of SAXS profiles*. *Nucleic Acids Res*, 2010. **38**(Web Server issue): p. W540-4.
250. *2.20 Angstrom resolution structure of 3-phosphoshikimate 1-carboxyvinyltransferase (AroA) from Coxiella burnetii*. <http://www.rcsb.org/structure/3ROI>.
251. *Crystal structure of a chorismate mutase/shikimate 5-dehydrogenase fusion protein from Clostridium acetobutylicum*. <http://www.rcsb.org/structure/3FBT>.
252. *Mycobacterium Tuberculosis Epsp Synthase in Unliganded State*. <http://www.rcsb.org/structure/2BJB>.
253. Bode, R., *[Gene-enzyme relationships of the arom aggregate of Schizosaccharomyces pombe]*. *Z Allg Mikrobiol*, 1983. **23**(4): p. 219-24.
254. Moore, R.B., et al., *A photosynthetic alveolate closely related to apicomplexan parasites*. *Nature*, 2008. **451**(7181): p. 959-963.
255. Hasan, N. and E.W. Nester, *Dehydroquinase synthase in Bacillus subtilis. An enzyme associated with chorismate synthase and flavin reductase*. *J Biol Chem*, 1978. **253**(14): p. 4999-5004.

Zusammenfassung

Multifunktionale Enzyme sind natürliche Fusionsproteine mit mehreren katalytischen Aktivitäten. Durch die räumliche Nähe der katalytischen Domänen profitieren diese Enzyme von der lokalen Konzentration der Reaktanten und können zusätzliche Besonderheiten wie allosterische Regulation und Substratkanalisierung aufweisen. Während Struktur und Funktion von monofunktionalen Enzymen im Allgemeinen sehr gut verstanden sind, sind Studien zu multifunktionalen Enzymen und zu Komplexen monofunktionaler Enzyme eher rar. Der Shikimatweg, welcher in sieben Schritten zu aromatischen Verbindungen führt, enthält eine Reihe bi-, tri-, tetra- und pentafunktionaler Enzyme in unterschiedlichen Organismen. Am bemerkenswertesten ist hier der pentafunktionale AROM-Komplex, welcher die zentralen fünf Schritte katalysiert, und dessen Struktur über Dekaden hin nicht aufgeklärt werden konnte.

In dieser Arbeit beleuchten wir diesen AROM-Komplex und die drei bifunktionalen Shikimatwegenzyme Tm_AroKB, Fp_AroEK und Mb_AroKE, indem wir sie sowohl funktionell als auch strukturell charakterisieren und mit ihren monofunktionalen Gegenstücken vergleichen. Für Tm_AroKB und Fp_AroEK erhielten wir katalytische Parameter, die mit denen der bekannten monofunktionalen Enzyme vergleichbar sind. Außerdem ist die katalytische Aktivität eines verkürzten, monofunktionalen Tm_AroK_{KB} vergleichbar mit der der entsprechenden Domäne des bifunktionalen Proteins. Interessanterweise konnten keine allosterischen Effekte zwischen den Domänen festgestellt werden. Die Analyse der K-Domäne von Mb_AroKE zeigte eine unerwartet niedrige Aktivität, welche deutlich unter der Aktivität der anderen untersuchten Proteine liegt. Das Genom des zugehörigen Organismus, *Methanoregula boonei*, enthält jedoch noch ein weiteres monofunktionales AroK-Homolog, welches möglicherweise die Hauptfunktion im Shikimatweg übernimmt.

In den Strukturen der bifunktionalen Proteine sind die einzelnen Domänen rigide miteinander verbunden, wobei alle aktiven Zentren zugänglich sind. Tm_AroKB ist ein Dimer und enthält bemerkenswerterweise ein permanent intrinsisch gebundenes NAD-Molekül. Die Strukturen der einzelnen Domänen aller charakterisierten bifunktionalen Proteine entsprechen im Wesentlichen denen ihrer monofunktionalen Gegenstücke, jedoch ist die räumliche Domänenanordnung bei

Fp_AroEK anders als bei Mb_AroKE. In beiden Anordnungen sind die aktiven Zentren jedoch nah beieinander. Es ist somit denkbar, dass diese Proteine Teil eines übergeordneten Komplexes sind, welcher den katalytischen Durchsatz des Shikimatwegs steigern könnte. Aus diesem Grund wurden im nächsten Schritt *in-silico*-Modelle solcher Komplexe konstruiert. Diese basieren auf der unvoreingenommenen Überlagerung der Strukturen von Fp_AroEK und Mb_AroKE auf die von Tm_AroKB und At_AroDE ohne weitere Modifikationen. Beide Modelle stellen trotz ihrer Unterschiede plausible, kompakte Komplexe mit zugänglichen aktiven Zentren dar.

Diese haben wir dann mit dem pentafunktionalen AROM-Komplex verglichen. In kinetischen Experimenten zeigte sich ein signifikant höherer Durchsatz als bei monofunktionalen *E. coli* Enzymen. Die Kristallstruktur des AROM-Komplexes zeigt eine sehr kompakte Anordnung von rigide miteinander verbundenen Domänen, welche sich jedoch von den *in-silico*-Modellen unterscheidet. Mittels SAXS- und XL-MS-Messungen konnte verifiziert werden, dass die Konformation des AROM-Komplexes in Lösung der Konformation im Kristall im Wesentlichen sehr ähnlich ist. Um Einblicke in die Konformationsdynamik des AROM-Komplexes zu gewinnen, haben wir die unterschiedlichen konformationellen Zustände der einzelnen enzymatischen Domänen aus der pdb-Datenbank auf die AROM-Kristallstruktur übertragen und somit ein strukturelles Ensemble erstellt, welches den konformationellen Raum des AROM-Komplexes widerspiegelt. Dieses Ensemble zeigt, dass der AROM-Komplex hinsichtlich der räumlichen Kompatibilität seiner Domänen optimiert ist. Ohne jegliche Koordination können die einzelnen Domänen ihre katalytisch notwendigen Konformationsänderungen durchlaufen und stoßen dabei nicht mit anderen Domänen zusammen.

Da der Shikimatweg nicht in Säugetieren vorkommt, ist er ein klassisches pharmakologisches Angriffsziel. Bis jetzt ist Glyphosat das einzige Herbizid, das im Shikimatweg angreift. Die Erkenntnisse dieser Arbeit zeigen neue Angriffspunkte zur Wirkstoffentwicklung auf. Dies betrifft insbesondere den AROM-Komplex, welcher essentiell in Pilzen und Protisten ist. Mit dem Ziel, katalytisch notwendige Konformationsänderungen zu inhibieren, könnte der konformationelle Raum des AROM-Komplexes gezielt angegriffen werden, was zur Entwicklung von neuartigen und spezifischen Fungiziden führen könnte.

Contributions

Marcus Hartmann proposed and defined the bifunctional enzymes projects presented in this work. Together with him, I made the CLANS map, for which Vikram Alva helped with the sequence analysis. The bifunctional enzymes presented here, were selected by me and Marcus.

For the AroKB project, Ulrich Koenninger cloned Tm_AroKB and Tm_AroK_{KB}. Reinhard Albrecht cloned Cf_AroKB, Am_AroKB, Gb_AroKB, and Of_AroKB. Maria Logotheti expressed Cf_AroKB, Am_AroKB, Gb_AroKB, and Of_AroKB. I cloned Tm_AroKB_SB. I expressed, purified, characterized and performed kinetic analysis of Tm_AroKB, Tm_AroK_{KB} and Tm_AroKB_SB and solved the crystal structures of Tm_AroKB, Tm_AroKB in complex with S3P and ADP, and Tm_AroKB_SB. Stefan Grüner performed the light scattering experiment for Tm_AroKB. Johannes Madlung, Mirita Franz-Wachtel, Boris Macek (Uni Tübingen), performed the mass spectrometry experiments.

For the AroEK and AroKE project, I cloned, expressed, and purified, Eb_AroEK, Fp_AroEK, Mb_AroKE and Mp_AroKE. Stefan Grüner performed the light scattering experiments for Fp_AroEK and Mb_AroKE. I characterized and performed kinetic analysis and solved the crystal structures of Fp_AroEK and Mp_AroKE. I further designed the *in silico* models based on Fp_AroEK and Mp_AroKE.

The AROM complex project was a collective idea that emerged during Marcus' and my discussions. Reinhard Albrecht cloned Ct_AROM and Th_AROM. In addition, he cloned expressed and purified the *E. coli* shikimate pathway enzymes. I expressed, purified and characterized Ct_AROM and Th_AROM. Matthias Flötenmeyer prepared the electron micrographs. I performed kinetic analysis of Ct_AROM. Alexander Leitner (ETH Zurich) conducted the XL-MS experiments and analyzed the data. Marcus Hartmann solved the crystal structure of AROM complex. I performed SAXS experiments for Ct_AROM and Th_AROM for which I received help from Eugene Valkov and beamline scientists Robert Rambo and Nathan Cowieson. Marcus and I analyzed and interpreted the SAXS data. Hongbo Zhu made the movie of the conformational space of AROM complex and majorly contributed in writing the methods for the same.

Additionally, Reinhard Albrecht and Kerstin Bär set-up all the crystallization screens. I froze and mounted most of the crystals. I further wrote this manuscript and prepared all the figures in it, for which I received constant advice from Marcus Hartmann and Reinhard Albrecht.

Acknowledgements

To begin with, I would like to thank Dr. Marcus Hartmann, for introducing me to the project, being my supervisor and letting me work in his group, during this PhD. I wish to thank him for the scientific guidance, for all his time, and discussions. Without his help and constant support, this work would not have been possible.

I am grateful to Prof. Dr. Andrei Lupas, for believing in my personal skills and letting me work in this department.

A very special gratitude goes to my TAC members, Prof. Dr. Karl Forchhammer, Prof. Dr. Birte Hoecker, Prof. Dr. Volkmar Braun, and Prof. Dr. Remco Sprangers, for their time and support. In addition, I would already like to thank the members of the examination board, Prof. Dr. Karl Forchhammer, Prof. Dr. Volkmar Braun, Prof. Dr. Andrei Lupas, and Prof. Dr. Thilo Stehle.

My sincere thanks to Dr. Reinhard Albrecht, for the discussions, his teaching skills, and helping with the experiments in the lab.

A special thanks to Dr. Tjeerd Dijkstra for carefully reading the draft of this thesis and his always-smiling nature.

I want to express my gratitude to Dr. Vikram Alva for the sequence analysis; Dr. Reinhard Albrecht and Kerstin Bär for the help in the crystallization of proteins; Dr. Hongbo Zhu for making the video of conformational space of AROM; Dr. Birte Hernandez-Alvarez for discussions, Dr. Matthias Flötenmeyer for the EM pictures; Stefan Grüner for help with LS experiments; Johannes Madlung, Mirita Franz-Wachtel and Boris Macek (University of Tübingen), for MS measurements; Prof. Dr. Alexander Leitner (ETH Zurich) for XL-MS experiments; Maria Logotheti for experimental collaboration.

My sincere gratitude goes to the people who helped with the German translation, Dr. Marcus Hartmann, Dr. Reinhard Albrecht, Dr. Birte Hernandez-Alvarez and Dr. Adrian Fuchs.

I would like to thank my friend, Ioanna Karamichali for always being there for me and I want to extend my gratitude to Dr. Edgardo Sepulveda, for his help during all these years.

I would also like to all thank all the present and past members of the Department (in alphabetical order), Jyoti Adlakha, Emilija Basina, Jens Baßler, Claire Bedez, Iuliia Boichenko, Philippe Chatelain, Manish Chaubey, Murray Coles, Silvia Deiss, Philipp Doerner, Mohammad ElGamacy, Hadeer Elhabashy, Dara Forouzan, Felix Gabler, Christopher Heim, Eva Hertle, Anja Heubach, Anita Jeganantham, Oliver Kohlbacher, Mateusz Korycinski, Jonas Kübler, Andriko von Kugelgen, Amit Kumar, Karin Lehmann, Alessia LoFaso, Marko Lozajic, Lorena Maldoner, Jörg Martin, Mikel Martinez, Ina Meuskens, Fabian Moertter, Seung-Zin Nam, Eugen Netz, Joana Pereira, David Rau, Franka Scharfenberg, Marc Schmollack, Martin Schueckel, Fabian Springer, Andrew Stephens, Shirley Tremel, Astrid Ursinus, Ivana Vujkovic Bukvin, Laura Weidmann, Matthias Wojtynek, Silvia Wuertenberger, Lukas Zimmermann and of course, everyone I forgot here.

During my time here, I have attended many courses, organized various events, and took part in many other activities, in which countless people have taught and helped me at every step. A huge thank you to all of you!

In addition, I want to thank my friends and family back home who always listened to my complaints patiently. Finally, I want to thank my parents, sister and Csaba for their constant support and motivation through years.

Last but not the least, to all the people who hindered my path, thank you for making me stronger.

Appendix

A.1 Sequences

A.1.1 *Thermotoga maritima* shikimate kinase and dehydroquinase synthase (Tm_AroKB)

Cloned in pet 22b

A.1.1.1 Primers

Forw: GAGAGA CATATG AGAATCTTCCTCGTTGG

Rev: GAGAGA AAGCTT ATCTACCACCTCCAGAAGC

A.1.1.2 Restriction sites

NdeI/HindIII

A.1.1.3 DNA sequence

```
atg aga atc ttc ctc gtt gga atg atg ggt tct gga aag agc act atc ggt aaa agg att
tcc gag gtg ctc gac ctt cag ttc ata gat atg gat gaa gag ata gaa aga aga gag gga
aga agc gtt cga agg att ttc gaa gaa gac ggt gag gag tac ttc cgc ctg aaa gaa aaa
gaa ctt ctt aaa gaa ctt gtg gaa aga gac aac gtg gtt gtg gca acg ggt gga ggt gtt
gtg gtc gat cca gag aac aga gag ctt ttg aag aag gaa aag act ctc ttt ctc tac gct
ccc cct gaa gtg ttg atg gaa aga gta aca aca gag aac aga cct ctt ctg agc gaa gga
aaa gag aga ata cgg gag atc tgg gaa aag aga aaa cag ttc tac gcg gag ttc aga agg
atc gac acc tcc agg ttg aac gag tgg gaa aca acc gca ctc gtt gtg ctg gag gct ctg
gac gag aaa gaa atc tca acg ata gaa aaa cca cac ctg gtg aag atc atc ctc ggt ggc
ttc aag agg gtg aga aac gaa gag ctg gtt ttc acc acg gag agg gtg gag aag ata tac
gga agg tac ctc ccg gag aat cgg ctt ctt ttt ccg gat gga gag gaa gtg aag acg ctg
gag cat gtc tcc aga gcg tac tac gaa ctc atc cga atg gac ttt ccc agg gga aag acc
ata gcg ggt gtc ggg gga ggt gct ctc acc gac ttc acc ggc ttt gta gcg agc acg ttt
aaa aga gga gtg gga ctt tct ttc tat cca aca aca ctt ctg gct cag gtg gac gct tcc
ggt ggt gga aag aat gcc atc gat ttc gct gga gtg aaa aac gtc gtt ggg act ttc aga
atg cca gac tac gtg atc ata gat ccc acc gtc acg ctt tcc atg gat gag ggc agg ttc
gaa gag gga gtc gtg gaa gcc ttc aag atg acg att ctt tcg ggt cgc ggg gta gaa ctc
ttc gat gag ccg gag aag att gag aag aga aat ctc aga gtt ctc agc gag atg gta aaa
atc tcc gtt gaa gag aaa gcg agg ata gtg atg gaa gat ccc tac gac atg ggt ttg agg
cac gcc ctg aat ctg gga cac aca ctc ggt cat gtg tac gag atg ctg gaa ggg gta cct
cac ggt ata gca gta gca tgg ggc atc gaa aaa gag acg atg tac ctg tac aga aaa gga
ata gtg cct aag gaa acc atg aga tgg atc gta gaa aag gtc aaa cag atc gta cca att
cct gtt cca tcc gtc gat gtt gag aaa gcc aga aat ctc att ctg aac gac aag aag atc
ctg aaa ggt tcc aga gtc agg ctt cct tac gtg aaa gaa atc gga aag atc gaa ttc ttg
gag gtc gat ccg ctc gag ctt ctg gag gtg gta ga
```

A.1.1.4 Amino acid sequence

MRIFLVGMMG SGKSTIGKRI SEVLDLQFID MDEEIERREG RSVRRIFEED GEEYFRLKEK ELLKELVERD
NVVVATGGGV VVDPENRELL KKEKTLFLYA PPEVLMERVT TENRPLLSEG KERIREIWEK RKQFYAEFRR
IDTSRLNEWE TTALVVLEAL DEKEISTIEK PHLVKIILGG FKRVRNEELV FTTERVEKIY GRYPENRLL
FPDGEEVKTL EHVSRAYYEL IRMDFPRGKT IAGVGGGALT DFTGFVASTF KRGVGLSFYP TTLAQVDAS
VGGKNAIDFA GVKNVVGTFR MPDYVIIDPT VTLSMDEGRF EEGVVEAFKM TILSGRGVEL FDEPEKIEKR
NLRVLSEMVK ISVEEKARIV MEDPYDMGLR HALNLGHTLG HVYEMLEGVP HGIAVAWGIE KETMYLYRKG
IVPKETMRWI VEKVKQIVPI PVPSVDVEKA RNLILNDKKI LKGSRVRLPY VKEIGKIEFL EVDPLELLEV
VLEHHHHHH

Cloned by: Reinhard Albrecht

A.1.2. *Thermotoga maritima* shikimate kinase cloned in pET His-TEV-Nde-GFP (Tm_AroK_{KB})

A.1.2.1 Primers

Forw: GAGAGA CATATG AGAATCTTCCTCGTTGG

REV: CTCTCT GGATCC TCAAGCCTCCAGCACAAACG

A.1.2.2 Restriction sites

NdeI/BamHI

A.1.2.3 DNA sequence

atg aga atc ttc ctc gtt gga atg atg ggt tct gga aag agc act atc ggt aaa agg att
tcc gag gtg ctc gac ctt cag ttc ata gat atg gat gaa gag ata gaa aga aga gag gga
aga agc gtt cga agg att ttc gaa gaa gac ggt gag gag tac ttc cgc ctg aaa gaa aaa
gaa ctt ctt aaa gaa ctt gtg gaa aga gac aac gtg gtt gtg gca acg ggt gga ggt gtt
gtg gtc gat cca gag aac aga gag ctt ttg aag aag gaa aag act ctc ttt ctc tac gct
ccc cct gaa gtg ttg atg gaa aga gta aca aca gag aac aga cct ctt ctg agc gaa gga
aaa gag aga ata cgg gag atc tgg gaa aag aga aaa cag ttc tac gcg gag ttc aga agg
atc gac acc tcc agg ttg aac gag tgg gaa aca acc gca ctc gtt gtg ctg gag gct tga

A.1.2.4 Amino acid sequence

MRIFLVGMMG SGKSTIGKRI SEVLDLQFID MDEEIERREG RSVRRIFEED GEEYFRLKEK ELLKELVERD
NVVVATGGGV VVDPENRELL KKEKTLFLYA PPEVLMERVT TENRPLLSEG KERIREIWEK RKQFYAEFRR
IDTSRLNEWE TTALVVLEA

Cloned by: Ulrich Könninger

A.1.3 *Thermotoga maritima* shikimate kinase and dehydroquinase synthase salt bridge mutant (Tm_AroKB_{SB})

Cloned in pet 22b

A.1.3.1 Primers

R 195 to A

Forw: P-GCG GTGGA GAAGATATACGGAAGG

Rev: CTCCGTGGTGAAAACCAG

D315 to A

Forw: P-CTGA GGCAGGTTC GAAGAG

Rev: CCATGGAAAGCGTGACGGTG

A.1.3.2 Restriction sites

NdeI/HindIII

A.1.3.3 DNA sequence

```
atg aga atc ttc ctc gtt gga atg atg ggt tct gga aag agc act atc ggt aaa agg att
tcc gag gtg ctc gac ctt cag ttc ata gat atg gat gaa gag ata gaa aga aga gag gga
aga agc gtt cga agg att ttc gaa gaa gac ggt gag gag tac ttc cgc ctg aaa gaa aaa
gaa ctt ctt aaa gaa ctt gtg gaa aga gac aac gtg gtt gtg gca acg ggt gga ggt gtt
gtg gtc gat cca gag aac aga gag ctt ttg aag aag gaa aag act ctc ttt ctc tac gct
ccc cct gaa gtg ttg atg gaa aga gta aca aca gag aac aga cct ctt ctg agc gaa gga
aaa gag aga ata cgg gag atc tgg gaa aag aga aaa cag ttc tac gcg gag ttc aga agg
atc gac acc tcc agg ttg aac gag tgg gaa aca acc gca ctc gtt gtg ctg gag gct ctg
gac gag aaa gaa atc tca acg ata gaa aaa cca cac ctg gtg aag atc atc ctc ggt ggc
ttc aag agg gtg aga aac gaa gag ctg gtt ttc acc acg gag gcg gtg gag aag ata tac
gga agg tac ctc ccg gag aat ccg ctt ctt ttt ccg gat gga gag gaa gtg aag acg ctg
gag cat gtc tcc aga gcg tac tac gaa ctc atc cga atg gac ttt ccc agg gga aag acc
ata gcg ggt gtc ggg gga ggt gct ctc acc gac ttc acc ggc ttt gta gcg agc acg ttt
aaa aga gga gtg gga ctt tct ttc tat cca aca aca ctt ctg gct cag gtg gac gct tcc
ggt ggt gga aag aat gcc atc gat ttc gct gga gtg aaa aac gtc gtt ggg act ttc aga
atg cca gac tac gtg atc ata gat ccc acc gtc acg ctt tcc atg gct gag ggc agg ttc
gaa gag gga gtc gtg gaa gcc ttc aag atg acg att ctt tcg ggt cgc ggg gta gaa ctc
ttc gat gag ccg gag aag att gag aag aga aat ctc aga gtt ctc agc gag atg gta aaa
atc tcc gtt gaa gag aaa gcg agg ata gtg atg gaa gat ccc tac gac atg ggt ttg agg
cac gcc ctg aat ctg gga cac aca ctc ggt cat gtg tac gag atg ctg gaa ggg gta cct
cac ggt ata gca gta gca tgg ggc atc gaa aaa gag acg atg tac ctg tac aga aaa gga
ata gtg cct aag gaa acc atg aga tgg atc gta gaa aag gtc aaa cag atc gta cca att
cct gtt cca tcc gtc gat gtt gag aaa gcc aga aat ctc att ctg aac gac aag aag atc
ctg aaa ggt tcc aga gtc agg ctt cct tac gtg aaa gaa atc gga aag atc gaa ttc ttg
gag gtc gat ccg ctc gag ctt ctg gag gtg gta gat
```


A.1.3.4 Amino acid sequence

MRIFLVGMMG SGKSTIGKRI SEVLDLQFID MDEEIERREG RSVRRIFEED GEEYFRLKEK ELLKELVERD
NVVVATGGGV VVDPENRELL KKEKTLFLYA PPEVLMERVT TENRPLLSEG KERIREIWEK RKQFYAEFRR
IDTSRLNEW ETTALVVLEAL DEKEISTIEK PHLVKIILGG FKRVRNEELV FTTEAVEKIY GRYLPENRLL
FPDGEEVKTL EHVSRAYYEL IRMDFPRGKT IAGVGGGALT DFTGFVASTF KRGVGLSFYP TTLAQVDAS
VGGKNAIDFA GVKNVVGTFR MPDYVIIDPT VTL SMAEGRF EEGVVEAFKM TILSGRGVEL FDEPEKIEKR
NLRVLSEMVK ISVEEKARIV MEDPYDMGLR HALNLGHTLG HVYEMLEGVP HGI AVAWGIE KETMYLYRKG
IVPKETMRWI VEKVKQIVPI PVPSVDVEKA RNLILNDKKI LKGSRVRLPY VKEIGKIEFL EVDPLELLEV
VDLEHHHHHH

Cloned by: Harshul Arora

A.1.4 *Feacalibacterium prausnitzii* shikimate dehydrogenase and kinase (FP_AroEK)

Codon optimized for expression in *E. coli* from Eurofins and cloned in pet22b

A.1.4.1 Primers

Forw: GAGAGA CATATG GAA TAC GGA CTC ATC GG

Rev: GAGAGA CTCGAG TTA ATG ACT AAA AAT TTC ATC CAG CG

A.1.4.2 Restriction sites

NdeI/XhoI

A.1.4.3 DNA Sequence

catatg atg gaa tat ggg ctg att ggt ggc cgg ctg ggt cac tcg tat agc aag gtc att
cac gag atg ttg tgc ggc tat cgc tat gac ctt tgt ccg ttg ccg acg gaa gag gaa gtc
cgt gct ttt ctc aca cgt cgc cag ttt cgg gcc atc aac gtg aca att ccc tac aaa ctg
ggt gtg atg gag tat tgc agc tac atc gat cct cat gcg aaa gcg att aat gcc gtg aac
acg atc gtg aat cgc aac ggt tta ctg tat ggc tac aat acc gac tat ccg ggt ttc tcc
tat ctg tgc gat gca cat ggg gtg gag ttc aaa gac cgc acg gta ctg att tta ggg aca
ggc ggg acc cac aac acc acg tgg gcc gta gcg cat gat cgt ggg gca aag cag att tac
acc gtt agt cgc cat ccc gat ccg gaa aag ggt gaa tta acc tat gca cag gcc ctg acc
aca ggt gcg cag att atc atc aat acg act cct gtg ggt atg tac ccg aat gcg gga gta
tct gcg ttg gat atc acc agc atg ccg ggc ctc gaa gcg gtt att gac gtg att tac aat
ccg gac aaa acc gaa ctt atc ctg cgt gct gaa gag ctg ggt gtc cca gtt gcc gtg ggt
ggt ctg gaa atg ctg gta gct caa gcg gtt tat gca gcc gaa ttc ttt ctg gat cgc aaa
ttc gaa gat gca ggc gcc gaa att gcc cgt gtg acg tcc gag ctg cgc cgt gag cag ttg
aac att gcg ctg att ggc atg ccg tca tcg ggc aaa agc acc tta ggc cgt gca tta gcc
gaa cgc tta ggt aaa cgc ttt gta gat ctg gac gag gaa att gtc aag gca gat ggc cgt
tcg att ccc gac atc ttt gcg gct gaa ggc gaa gca ggc ttc cgc aaa ctg gaa gcc gcg
caa act gcg cgg ttt gcg cgt gag aat cgc cag gtg att tca tgt gga ggc gga gtt gtg
aaa gat cca gct aac ctt cgt gcg ctg cat gct aac ggc att gtc ctg ttc atc gat cgt
ccg ctg gaa gat ctg ctc gtc ggc ggt gga cgc cca ctc agt act tct ccg gaa gcc ctg
aaa acc atg gaa gcg caa cgc cgc cca ctt tac ctg gct gcc gct gat gca gtt atc ccg
aac aaa acg acc cct gca gat gcg gtt act gcg gcg atg gaa gcc ttg gac gag atc ttt
agc cac ctcgag

A.1.4.4 Protein Sequence

```
MMEYGLIGGR LGHSYSKVIH EMLCGYRYDL CPLPTEEEVR AFLTRRQFRA INVTIPYKLV VMEYCSYIDP
HAKAINAVNT IVNRNGLLYG YNTDYPGFSY LCDAHGVEFK DRTVLILGTG GTHNTTWAVA HDRGAKQIYT
VSRHPDPEKG ELTYAQALTT GAQIIINTTP VGMYPNAGVS ALDITSMPGL EAVIDVIYNP DKTELILRAE
ELGVPVAVGG LEMLVAQAVY AAFFFLDRKF EDAGAEIARV TSELRREQLN IALIGMPSSG KSTLGRALAE
RLGKRFVDLD EEIVKADGRS IPDIFAAEGE AGFRKLEAAQ TARFARENRO VISCGGGVVK DPANLRALHA
NGIVLFIDRP LEDLLVGGGR PLSTSPEALK TMEAQRRLY LAAADAVIPN KTTPADAVTA AMEALDEIFS
HLELEHHHHHH
```

A.1.5 *Methanoregula boonei* shikimate dehydrogenase and kinase (Mb_AroKE)

Codon optimized for expression in *E. coli* from Eurofins and cloned in pet22b

A.1.5.1 Primers

Forw: GAGAGACATATGGGCGGCAGCGGCATGAAGGTCGTCCTTACCGC

Rev: GAGAGAGGATCCTCATGCAAGCATCTCCCG

A.1.5.2 Restriction sites

NdeI/XhoI

A.1.5.3 DNA Sequence

```
ggc ggt tct ggt atg aaa cgc att gta ctg ttt ggc tat cgc ggt act ggc aaa act gcg
atc ggt acc gtt ctg gcg cag aaa ctg ggc gtt ccg ttt ctg gat acg gat gct ctg gtt
gaa cag caa gct gga cgc acg atc ccg gaa atc ttt cgc gat tcc ggt gaa gcc gga ttt
cgc gct cgc gaa cgt gaa gcg gtg tct ggt ctg ccg gat cgt gat gcg att atc gcg aca
ggt ggc ggt gtc gtc atg gac ccg gcg aac atg gaa cat ctg cgg aaa gaa agc gtg tgt
gtc ctg tta tcg gcg gat ccg aat gtc atc gga cat cgc ctg gcc cat gcc cca cgt ccg
gct ttg acg agc ctc tca ccc acg gat gag att acg gcg atg ctc aaa cac cgc cgc cct
gct tac gct gca gcc gca gat ttc tgc att gat aca ggc cgt act acc gcc ggc gaa gcc
gcc gag aaa atc ctt acg tta ctt ggc gcg ggt agc att ccg gat acc gcc cgt cac acc
gca gcc cgc tgg ttt gca gca acg cca ctt cct gcg ccg gaa aag gaa gaa ctg gaa cgc
aaa ttg ctg ggt ccg ggg tat gat ccg cag act cgc ttc ctg ggc gtt gcc ggc tgg ccg
tgt ggt cat tcg aaa agt cct gtg ctc ttc aac cgt ttg ttt gag cac tac cgc ctc aat
tgc cac tat acc cgt ttc gaa gcc ccg gaa att ggc cca gtc atg gaa atg gcg cgt ctt
att tcg gct aaa ggg tta agc gta acc att ccc ttc aaa cag gac gtg atg tcc tac ctg
gac gag att gat gaa gca gcg cag aaa att ggg gca gta aac acc gtg gtg ttt gcg tgt
ggc cgt gcg tat ggg tgg aac acc gat tgg atc ggg att cgt aag cct ctg gcc cat ctg
agt ggc tca cgg gcg gtt ttg ttg ggc gcc ggt ggg gta gct gca gca gcg gcg tat gcg
ctg cgt gac ctg gat atg gag atc atc atc ctg aat cgc aca ccc gag aag gca cgg gcg
tta gct gaa cgc aca ggt tgt cgc ttt ggt gcc tgg gac gat ttt gac cgc acg aat ccg
gat ctg gtt gtt aac gca acc agc att ggc atg caa ccc gat act gga agt ccg tta cgt
gac gac cag tta aag aaa gag atg acc gtg tgc gac ttg gtg tat act cca ccg gtc acc
tcc ctg atc gca gcc gcg cgt aaa gcg ggc tgc acc aca att ctg ggc acc gaa acc ttc
gtg tac caa gca caa gag cag ttt cgc ctg ttc ttc gga att gat gtg cca gac acg acc
att cgc gaa att ctg gcc
```

A.1.5.4 Protein sequence

```
GGSGMKRIVL FGYRGTGKTA IGTVLAQKLG VPFLDLDALV EQQAGRTIPE IFRDSGEAGF RAREREAVSG
LPDRDAIIAT GGGVVMDPAN MEHLRKESVC VLLSADPNVI GHRLAHAPRP ALTSLSPTDE ITAMLKHRRP
AYAAAADFCI DTGRTTAGEA AEKILTLLGA GSIPDTARHT AARWFAATPL PAPEKEELER KLLGPGYDPQ
TRFLGVAGWP CGHSKSPVLF NRLFEHYRLN CHYTRFEAPE IGPVMEMARL ISAKGLSVTI PFKQDVMSYL
DEIDEAAQKI GAVNTVVFAC GRAYGWNTDW IGIRKPLAHL SGSRAVLLGA GGVA AAAAYA LRDLDMEIII
LNRTPEKARA LAERTGCRFG AWDDFDR TNP DLVVNATSIG MQPDTGSPLR DDQLKKEMTV CDLVYTPPVT
SLIAAARKAG CTTILGTETF VYQAQEQFRL FFGIDVPD TT IREILALEHHHHHH
```

Cloned by: Harshul Arora

A.1.6. *Chaetomium thermophilum* AROM complex

cloned in pET28b

A.1.6.1 Primers

Forw: GAGAGAGAGACCATGGCCACC

Rev: CACGTTACAGCTCGAGTTCAAC

A.1.6.2 Restriction sites

NcoI and XhoI

A.1.6.3 Gene Sequence

```
gag aga gag acc atg gcc acc gca aat gtt gcc ggt gcc ggt agc ggt agc gaa ccg
acc cgt att gca att ctg ggt aaa gaa gat att atc gtc gat cat ggc atc tgg ctg aat
ttt gtt gca cat gat ctg ctg caa acc ctg ccg agc agc acc tat gtt ctg att acc gat
acc aat ctg tat acc acc tat gtg cca ccg ttt cag gca gtt ttt gaa gca gca gca ccg
cgt gat gtt cgt ctg ctg acc tat gca att ccg cct ggt gaa tat agc aaa agc cgt gaa
acc aaa gcc gaa att gaa gat tgg atg ctg agc cat gca tgt acc cgt gat acc gtt att
att gca ctg ggt ggt ggt gtt att ggt gat atg att ggt tat gtt gca gcc acc ttt atg
cgt ggt gtt cgt ttt gtt cag gtt ccg acc acc ctg ctg gca atg gtt gat agc agc att
ggt ggt aaa acc gca att gat acc ccg atg ggt aaa aat ctg att ggt gca ttt tgg cag
cct cgt cgt atc tat att gat ctg gca ttt ctg gaa aca ctg ccg gtt cgt gaa ttt atc
aat ggt atg gca gaa gtg att aaa acg gca gcc att tgg aat gaa acc gaa ttt acc gca
ctg gaa gaa aat gca gca gcg att ctg gaa gca gtt cgt agc aaa gca agc agt ccg gca
gca cgt ctg gca ccg att cgt cat att ctg aaa cgt att gtt ctg ggt agc gca cgt gtt
aaa gcc gaa gtt gtt agc gca gat gaa cgt gaa ggt ggt ctg cgt aat ctg ctg aac ttt
ggt cat agc att ggt cat gcc tat gaa gca atc ctg gca ccg cag gtt ctg cat ggt gaa
tgt gtt gca att ggt atg gtt aaa gaa gca gaa ctg gca cgt tat ctg ggt gtt ctg cgt
ccg agc gca gtt gca cgt ctg acc aaa ctg att gca agc tat gac ctg ccg acc agc gtt
cat gat aaa cgc att gca aaa ctg agc gca ggt aaa gaa tgt ccg gtt gat gta ctg ctg
caa aaa atg gcc gtt gac aaa aaa aac gag ggt cgc aaa aaa aaa att gtt ctg ctg agc
gca atc ggc aaa acc tat gaa aaa aaa gca acc gtt gtt gat gat cgt gca att cgt ctg
ggt ctg agc ccg agc gtt cgt gtt aca ccg ggt gtt ccg aaa ggt ctg agc gtt acc gtt
acc cct ccg ggt agc aaa agt att agc aat cgt gca ctg gtt ctg gca gcc ctg ggt gaa
ggc acc acc cgt att cat ggc ctg ctg cat agt gat gat gtt cag tat atg ctg gca gca
att gaa cag ctg cat ggc gca gat ttt tca tgg gaa gat gcc ggt gaa att ctg gtt gtt
acc ggc aaa ggt ggt aaa ctg caa gca agc aaa gaa ccg ctg tat ctg ggt aat gca ggc
acc gca agc cgt ttt ctg acc agc gtg gtt gca ctg tgt gca ccg agt gca gtt agc agt
```

acc gtt ctg acc ggc aat gca cgt atg aaa gtt cgt ccg att ggt gcc ctg gtt gat gcc
ctg cgt gca aat ggt gtt ggt gtt aaa tat ctg gaa aaa gag aaa agc ctg ccg gtt gaa
gtt gat gca gcc ggt ggt ttt gca ggc ggt gtg att gaa ctg gca gcc acc gtt agc agc
cag tat gtg agc agc att ctg atg gca gca ccg tat gca cat cag ccg gtt acc ctg cgt
ctg gtt ggt ggc aaa ccg att agc cag ccg tat att gat atg aca att gca atg atg gcc
agc ttt ggc att aaa gtt gaa cgt agc gca gaa gat ccg aac acc tat ctg atc ccg aaa
ggt gtg tat aaa aac cct ccg gaa tat gtg gtt gaa agt gat gca agc agc gca acc tat
ccg ctg gca gtt gca gca att acc ggc acc acc tgt acc att ccg aat att ggt agc gaa
agc ctg caa ggt gat gca cgt ttt gca gtt gaa gtg ctg cgt cct atg ggt tgt gcc gtt
gaa cag acc gca acc agc acc acc gtt aca ggt ccg cct att ggc acc ctg aaa gcc att
ccg cat gtt gat atg gaa ccg atg acc gat gca ttt ctg acc gca gcc gta ctg gca gca
gtt gcc gat ggt aca acc cag att aca ggt att gca aat cag cgt gtg aaa gaa tgt aat
cgt att gca gcc atg aaa gat cag ctg gca aaa ttt ggt gtt cag tgc aat gaa ctg gaa
gat ggc att gaa gtt att ggt aaa ccg tat caa gaa ctg cgt aat ccg gtt gaa ggt atc
tat tgc tat gat gat cac cgt gtt gca atg agc cat agc gtt ctg agc acc att agt ccg
cat ccg gtt ctg att ctg gaa cgt gaa tgt acc gca aaa acc tgg cct ggt tgg tgg gat
att ctg agc cag ttt ttc aaa gtt cag ctg gat ggt gaa gaa gat ccg acc aaa cgt acc
acc cag agc acc cag cag gtt cgt aaa ggc acc gat cgt agc att ttt atc gtt ggt atg
cgt ggt gca ggt aaa agc acc gca ggt cgt tgg atg agc gaa ctg ctg aaa cgt ccg ctg
gtt gat ctg gat gca gaa ctg gaa cgt cgt gaa ggt atg acc att ccg gaa att att cgt
ggt gaa cgt ggt tgg gaa ggt ttt cgt cag gca gag ctg gaa ctg ctg caa gat gtt atc
aaa aat cag agc aaa ggc tac atc ttt agc tgt ggt ggt ggt att gtt gaa acc gaa gca
gca cgt aaa ctg ctg atc gat tat cat aaa aat ggt ggt ccg gtg ctg ctg gtt cat cgt
gat acc gat cag gtt gtt gaa tat ctg atg cgt gat aaa acc cgt ccg gca tat agc gaa
aat att cgc gaa gtt tat gaa cgt cgt aaa ccg tgg ttt tat gaa tgc agc aat ctg caa
tat cat agt ccg cac gaa gat ggt agc gaa gcc ctg ctg caa cct ccg gca gat ttt gca
cgt ttt gtt aaa ctg att gca ggt cag agt acc cat ctg gaa gat gtt cgt gca aaa aaa
cac agc ttt ttt gtt agc ctg acc gtt ccg aat gtt gca gat gca ctg gat att att ccg
cgt gtt gtt gtt ggt agt gat gca gtt gaa ctg cgt gtt gat ctg ctg gaa agc tat gaa
ccg gaa ttt gtt gca cgt cag gtt gca ctg cgt gca gcc cag gtt ccg att gtt
tat acc gtt cgt acc cag agc cag ggt ggt ggt ggt aaa ttt cct gat gaa gat tat gat ctg gca
ctg cgt ctg tat cag aca ggt ctg cgt agc ggt gtg gaa tat ctg gat ctg gaa atg aca
atg ccg gat cat att ctg caa gca gtt acc gat gca aaa ggt ttt acc agc att att gca
agc cat cat gat ccg cag tgt aaa ctg agc tgg aaa agc ggt agc tgg att ccg ttt tat
aac aaa gca ctg caa tat ggc gac gtg att aaa ctg gtg ggt gtt gcc cgt gaa atg gca
gat aat ttt gca ctg acc aac ttt aaa gcc aaa atg ctg gca gcc cat gat aac aaa ccg
atg att gcc ctg aat atg ggt aca gcc ggt aaa ctg tca cgt gtt ctg aat ggt ttt ctg
act ccg gtt agc cat ccg gca ctg ccg agc aaa gcc gca ccg ggt cag ctg agc gca acc
gaa att cgc cag gca ctg agc ctg att ggt gaa att gaa ccg aaa tgc ttt tac ctg ttt
ggt aaa ccg att agc gca agc cgt agc cct gca ctg cat aat acc ctg ttt tac aaa acg
ggt ctg ccg cat cat tat agc cgt ttt gaa acg gat gaa gca agc aaa gcc ctg gaa agc
ctg att cgt agt ccg gat ttt ggt ggt gcc agc gtt acc att ccg ctg aaa ctg gac att
atg ccg ctg ctg gat agc gcc acc gat gca gcc cgt acc att ggt gca gtt aat acc att
att cct cag acc cgt gat ggt agc acc aca acc ctg gtt ggc gat aat acc gat tgg cgt
ggt atg gtt cat gcg ctg ctg cat agt agc ggt agc ggt tca gtt gtt cag cgt acc gca
gca ccg cgt ggt gcc gca atg gtt gtg ggt agc ggt ggc acc gca cgt gca gca att tat
gcc ctg cat gat ctg ggt ttt gca ccg att tgg att gtt gcg cgt agc gaa gaa cgt gtt
gcc gaa ctg gtt cgt ggt ttt gat ggc tat gat ctg cgt agc acc tct ccg cat cag
ggt aaa gat aat atg ccg agc gtt gtt att agc aca att ccg gca acc cag ccg att gat
ccg agc atg cgt gaa gtg att gtt gaa gtt ctg aaa cat ggt cat ccg agc gca gaa ggt
aaa gta ctg ctg gaa atg gca tat cag cct ccg cgt aca ccg ctg atg acc ctg gca gaa
gat cag ggt tgg cgt acc gtt ggt ggt ctg gaa gtt ctg gca gca cag ggt tgg tat cag
ttt caa ctg tgg acc ggc att act ccg ctg tat gaa gaa gca cgc gca gca gtt atg ggt
gaa gat agc gtt gaa ctc gag cgt gaa cgt g

A.1.6.4 Amino acid Sequence

```
MATANVAGAG GSGSEPTRIA ILGKEDIIVD HGIWLNFFVAH DLLQTLPSST YVLITDTNLY TTYVPPFQAV
FEAAAPRDVR LLTYAIPPGGE YSKSRETKAE IEDWMLSHAC TRDTVIIALG GGVIGDMIGY VAATFMRGVR
FVQVPPTLLA MVDSSIGGKT AIDTPMGKNL IGAFWQPRRI YIDLAFLETL PVREFINGMA EVIKTAAIWN
ETEFTALEEN AAAILEAVRS KASSPAARLA PIRHILKRIV LGSARVKAEV VSADEREGGL RNLLNFGHSI
GHAYEAILAP QVLHGECVAI GMVKEAELAR YLGVLRPSAV ARLTKLIASY DLPTSVHDKR IAKLSAGKEC
PVDVLLQKMA VDKKNEGRKK KIVLLSAIGK TYEKKATVVD DRAIRLVLSP SVRVTPGVPK GLSVTVTPPG
SKSISNRALV LAALGEGTTR IHGLLHSDDV QYMLAAIEQL HGADFSWEDA GEILVVTGKG GKLQASKEPL
YLGNAGTASR FLTSSVALCA PSAVSSTVLT GNARMKVRPI GALVDALRAN GVGVKYLEKE KSLPVEVDAA
GGFAGGVIEL AATVSSQYVS SILMAAPYAH QPVTLLRVGG KPISQPYIDM TIAMMASFGI KVERSAEDPN
TYLIPKGVYK NPPEYVVESD ASSATYPLAV AAITGTTCTI PNIGSESLQG DARFAVEVLR PMGCAVEQTA
TSTTVTGPPi GTLKAIPHVD MEPMTDAFLT AAVLAAVADG TTQITGIANQ RVKECNRIAA MKDQLAKFGV
QCNELEDGIE VIGKPYQELR NPVEGIYCYD DHRVAMSHSV LSTISPHVPL ILERECTAKT WPGWWDILSQ
FFKVQLDGEE DPTKRTTQST QQVRKGTDRS IFIVGMRGAG KSTAGRWMSE LLKRPLVDLD AELERREGMT
IPEIIRGERG WEGFRQAELE LLQDVIKQNS KGYIFSCGGG IVETEAAARKL LIDYHKNNGP VLLVHRDTDQ
VVEYLMRDKT RPAYSENIRE VYERRKPFY ECSNLQYHSP HEDGSEALLQ PPADFARFVK LIAGQSTHLE
DVRAKKHSFF VSLTVPNVAD ALDIIPRVVV GSDAVALRVD LLESYEPEFV ARQVALLRAA AQVPIVYTVR
TQSQGGKFPD EDYDLALRLY QTGLRSGVEY LDLEMTMPDH ILQAVTDAKG FTSI IASHHD PQCKLSWKSG
SWIPFYNKAL QYGDVIKLVG VAREMADNFA LTNFKAKMLA AHDNKPMIAL NMGTAGKLSR VLNGLFPTVS
HPALPSKAAP GQLSATEIRQ ALSLIGEIEP KSFYLFKPI SASRSPALHN TLFYKTGLPH HYSRFETDEA
SKALESLIRS PDFGGASVTI PLKLDIMPLL DSATDAARTI GAVNTIIPQT RDGSTTTLVG DNTDWRGMVH
ALLHSSSGSGS VVQRTAAPRG AAMVVGSGGT ARAAIYALHD LGFAPIWIVA RSEERVAELV RGFDDGYDLRR
MTSPHQGKDN MPSVVISTIP ATQPIDPSMR EVIVEVLKHG HPSAEGKVLL EMAYQPPRTP LMTLAEDQGW
RTVGGLEVLA AQGWYQFQLW TGITPLYEEA RAAVMGEDSV ELERERLEHHHHHH
```

Cloned by: Reinhard Albrecht

A.1.7. *Thielavia heterothallica* AROM complex cloned in pET28b with NcoI and XhoI

A.1.7.1 Primers

Forw: GAGAGAGAGACCATGGCCG

Rev: CTCTCTCTCTCTCGAGGCTGC

A.1.7.2 Restriction sites

NcoI and XhoI

A.1.7.3 Gene Sequence

```
gag aga gag acc atg gcc ggt aca agc agc agc ggt ccg acc cgt att gca att ctg ggt
aaa gaa gat att gtt gtc gat cat ggc atc tgg ctg aat ttt gtt acc cat gat ctg ctg
caa aat ctg ccg agc agc acc tat gtt ctg att acc gat acc aat ctg tat cac acc tat
gtg cca ccg ttt cag gca gtt ttt gaa gca gcc gca ccg aaa gat gtt cgt ctg ctg acc
tat gca att ccg cct ggt gaa tat agc aaa ggt cgt gaa acc aaa gcc gaa att gaa gat
tgg atg ctg agc cat acc tgt acc cgt gat acc gtt att att gca ctg ggt ggt ggt gtt
att ggt gat atg att ggt tat gtt gca gcc acc ttt atg cgt ggt gtt cgt ttt gtt cag
gtt ccg acc acc ctg ctg agc atg gtt gat agc agc att ggt ggt aaa acc gca att gat
gtt ccg atg ggt aaa aat ctg att ggt gca ttt tgg cag ccg aaa cgc att tat atc gat
ctg gca ttt ctg gaa acc ctg ccg gtt cgt gaa ttt atc aat ggt atg gca gaa gtt att
```

aaa acg gca gcc att tgg aat gaa gca gaa ttt acc gca ctg gaa gaa aac gca ccg gct
att ctg gaa gca att cgt agt aaa ggt agc agc ccg agc gca cgt ctg gca ccg att cgt
cat att ctg aaa cgt att gtt ctg ggt agt gcc ggt gtt aaa gca cag gtt gtt agc gca
gat gaa cgt gaa ggt ggt ctg cgt aat ctg ctg aac ttt ggt cat agc att ggt cat gcc
tat gaa gca atc ctg gca ccg cag gtt ctg cat ggt gaa tgt gtt gca att ggt atg gtt
aaa gaa gca gaa ctg gca cgt ttt ctg ggt gtt ctg cgt ccg gat gca gtt gca cgt ctg
acc aaa tgt att gca agc tat gac ctg ccg acc agc gtt cat gat aaa cgc att gca aaa
ctg agc gca agc aaa gaa tgt ccg gtt gat gtg ctg ctg caa aaa atg gca gtt gac aaa
aaa aac gaa ggc agc aaa aaa aaa atc gtt ctg ctg agc gcc att ggc aaa acc tat gaa
ccg aaa gca acc gtt gtt gat gat cgt gcc att cgt att gtg ctg agc ccg agc att cgt
gtt aca ccg ggt gtt ccg gaa aat ctg agc gtt agc gtt acc cct ccg ggt agc aaa agc
att agc aat cgt gca ctg att ctg gca gcc ctg ggt gaa ggc acc acc cgt att cat ggc
ctg ctg cat agt gat acc cag tat atg ctg acc gcg att gca cag ctg gaa ggt gca
acc tat acc tgg gaa gat gcc ggt gaa gtt ctg gtt gtt aaa ggt aaa ggt ggt aaa ctg
aaa gca agc gca gaa ccg ctg tat ctg ggt aat gca ggc acc gca agc cgt ttt ctg acc
agc gtg gtt gca ctg tgt agc ccg acc agt gtt agc agc acc gtt ctg acc ggc aat gca
cgt atg aaa gtt cgt ccg att ggt ccg ctg gtt gat gca ctg cgt gca aat ggt gtt ggt
gtt aaa tat ctg gaa aaa gag aaa agc ctg cct gtt cag gtg gat gca aat gca ggt ttt
gcg ggt ggt ctg att gaa ctg gca gcc acc gtt agc agc cag tat gtg agc agc att ctg
atg gct gca ccg tat gca cat aaa cct gtt acc ctg cgt ctg gtt ggt ggc aaa ccg att
agc cag ccg tat att gat atg acc att gcc atg atg gcc agc ttt ggt att aat gtt gaa
cgt agc agc gaa gat ccg aac acc tat cat att ccg cag ggt gtg tac aaa aat ccg gca
gat tat gtt gtt gaa agt gat gca agc agc gca acc tat ccg ctg gca gtt gca gca att
acc ggc acc acc tgt acc att ccg aat att ggt agc gca agc ctg caa ggt gat gca cgt
ttt gca gtt gaa gtg ctg cgt ccg atg ggc tgt acc gtt gaa cag acc gaa acc agc acc
acc gtt aca ggt ccg cct att ggc acc ctg aaa gcc att gaa cat gtt gat atg gaa ccg
atg acc gat gca ttt ctg acc gca agc gtg ctg gca gca gtt gca agc ggt aca acc cgc
att aca ggt att gca aat cag cgt gtt aaa gag tgt aat cgt att gca gcc atg aaa gat
cag ctg gca aaa ttt ggt gtg cat tgt aat gaa ctg gaa gat ggt att gaa gtt acc ggc
aaa ccg tat caa gag ctg caa aac ccg acc gat ggt gtt tat tgt tat gat gat cac cgt
gtt gcc atg agc ttt agt gtt ctg gca acc gca gca ccg cat cgt gtt ctg atc ctg gaa
cgt gaa tgt acc gca aaa acc tgg cct gga tgg tgg gat att ctg agc cag gca ttt aaa
gtt cat ctg gca ggt gaa gaa gat ccg acc aaa aaa cag gtt gtt cag ccg agc ggt agc
ggc acc gat cgt agc att ttt atc att ggt atg cgt ggt gca ggt aaa agc acc gca ggt
cgt tgg atg agc gaa att ctg caa cgt ccg ctg gtt gat ctg gat gtt gaa ctg gaa cgt
cgt gaa ggt atg acc att ccg gaa att att cgt ggt gaa cgt ggt tgg gaa ggt ttt cgt
aaa gca gag ctg gaa ctg ctg gaa gat gtt atc aaa aat cag agc aaa ggc tac atc ttt
agc tgt ggt ggt ggt att gtt gaa acc gaa gaa gca cgt aaa ctg ctg att agc tat cat
aaa agc ggt ggt tgt gtt ctg atg gtt cat cgt gat acc gat cag gtt gtt gaa tat ctg
atg cgt gat aaa acc cgt ccg gca tat agc gaa aat att cgc gaa gtt tat tat cgt cgc
aaa ccg tgg ttt gaa gaa tgc agc aat ttt cag tat cat agt ccg cat ctg gat ggt agc
gaa gca ctg cgt ggt ccg cct gtt gat ttt tca cgt ttt ctg agc gtt att tgt ggt cgt
agc cgt cat ctg gaa gaa gtg cgt aaa aaa aaa cac agc ttt ttt gtt agc ctg acc gtg
ccg aat att agc acc gca ctg gat att att ccg ggt gca gtt gtt ggt agt gat gca gtt
gaa ctg cgt gtt gat ctg ctg gaa agc tat gat ccg gaa ttt gtt gca cat cag gtt gca
ctg ctg cgt agc gtt gcc aaa att ccg att gtt tat acc gtt cgt acc gtt agc cag ggt
ggt cgt ttt ccg gat gat gat tat gaa ctg gca cgc cgt ctg tat cag att ggt ctg cgt
aca ggt gtg gaa ttt ctg gat ctg gaa gtt acc atg cct gcc gat att ctg gaa gca gtt
acc gtg agc aaa ggt ttt acc cgt att att gca agc cat cat gat ccg cag agc aaa ctg
agc tgg cgt aat ggt agc tgg att ccg ttt tat aac aaa gcc ctg caa tat ggc gac gtg
att aaa ctg gtt ggt atg gca ggc gaa att agc gat aat ttt gca ctg acc agc ttt aaa
acc aaa atg ctg gca gca cat gat acc ccg att att gcg ctg aat atg ggt ccg gca ggt
aaa ctg agc cgt gtg ctg aat ggt ttt atg act ccg gtt agc cat ccg gat ctg cct gca
aaa gca gca ccg ggt cag ctg agc gca acc gaa att cgt cag gca ctg gca ctg att ggc
gaa ctg gaa ccg aaa agt ttt tac ctg ttt ggt aaa ccg att agc agc agc cgt agt ccg
gca ctg cat aat acc ctg ttt ggc gca aca ggt ctg ccg cat cat tat agc ctg ttt gaa
acg gat aaa gcc agt gat gtt cag gat ctg att cgt gca ccg cat ttt ggc ggt gca agc

ggt acc att ccg ctg aaa ctg gac att gtt ccg ctg ctg gat gat att agc gaa gcc gca
 cgt gca att ggt gca gtt aat acc att att ccg att ccg agc gca gca aaa ggt gat aac
 aaa acc acc ctg ctg ggt gat aat acc gat tgg cgt ggt atg gtt ttt gcc ctg cgt agt
 gcc ggt gtt gca agc cct ccg ggt gcc ggt atg gtt gtg ggt agc ggt ggc acc acc cgt
 gca gcc att tat gcc ctg cat agc ctg ggt tat gca ccg att tat gtt gtt gca cgt aca
 ccg gaa cgt gtt aaa gaa ctg gca ggc ggt ttt ccg gca gat tat cgt att cag cat ctg
 gca acc cgt gaa gaa gca gtt gcg gtt gca gtt gca gca cgt cat gaa ctg ccg acc gtt
 gtg att agc acc att cct gcg gat aaa ccg atc gat agc ggt atg cgc gaa gtt ctg gtt
 acc gca ctg caa ggt ggt ggt gca gca ggt aaa gaa aaa ggt gaa ccg cgt gtg ctg ctg
 gaa atg gca tat aca ccg cgt cat aca ccg ctg atg cag ctg gca gaa gat gca ggt tgg
 gtt aca att ccg ggt ctg gaa gta ctg gca gcc cag ggt tgg tat cag ttt caa ctg tgg
 acc ggc att act ccg ctg tat gca gat gca cgt gcc gca gtt atg ggt aat gaa agc agc
 ctc gag aga gag aga g

A.1.7.4 Amino acid Sequence

MAGTSSSGPT RIAILGKEDI VVDHGIWLNF VTHDLLQNLP SSTYVLITDT NLYHTYVPPF QAVFEAAAPK
 DVRLLLTYAIP PGEYSKGRET KAEIEDWMLS HTCTRDTVII ALGGGVIGDM IGYVAATFMR GVRFVQVPTT
 LLSMVDSSIG GKTAIDVPMG KNLIGAFWQP KRIYIDLAFI ETLPVREFIN GMAEVIKTAI IWNEAEFTAL
 EENAPAILEA IRSKGSSPSA RLAPIRHILK RIVLGSAGVK AQVVSADERE GGLRNLLNFG HSIGHAYEAI
 LAPQVLHGEC VAIGMVKEAE LARFLGVLRP DAVARLTKCI ASYDLPTSVA DKRIAKLSAS KECVVDVLLQ
 KMAVDKKNEG SKKKIVLLSA IGKTYEPKAT VVDDRAIRIV LSPSIRVTPG VPENLSVSVT PPGSKSISNR
 ALILAALGEG TTRIHGLLHS DDTQYMLTAI AQLEGATYTW EDAGEVLVVK GKGGKLGASA EPLYLGNAGT
 ASRFLTSVVA LCSPTSVSST VLTGNARMKV RPIGPLVDAL RANGVGVKYL EKEKSLPVQV DANAGFAGGL
 IELAATVSSQ YVSSILMAAP YAHKPVTLRL VGGKPISQPY IDMTIAMMAS FGINVERSSE DPNTYHIPQG
 VYKNPADYVV ESDASSATYP LAVAAITGTT CTIPNIGSAS LQGDARFAVE VLRPMGCTVE QTETSTTVTG
 PPIGTLKAIE HVDMEPMTDA FLTASVLA AV ASGTTRITGI ANQRVKECNR IAAMKDQLAK FGVHCNELED
 GIEVTGKPYQ ELQNPTDGVY CYDDHRVAMS FSVLATAAPH RVLILERECT AKTWPGWWDI LSQAFKVHLA
 GEEDPTKKQV VQPSGSGTDR SIFIIGMRGA GKSTAGRWMS EILQRPLVDL DVELERREGM TIPEIIRGER
 GWEGFRKAEL ELLEDVIKNQ SKGYIFSCGG GIVETEEARK LLISYHKSGG CVLMVHRD TD QVVEYLMRDK
 TRPAYSENIR EVYYRRKPF EEC SNFQYHS PHLDGSEALR GPPVDFSRFL SVICGRSRHL EEVRKKKHSF
 FVSLTVPNIS TALDIIPGAV VGSDAVELRV DLLESYDPEF VAHQVALLRS VAKIPIVYTV RTVSQGRFP
 DDDYELARRL YQIGLRTGVE FLDLEVTMPA DILEAVTVSK GFTRIIASHH DPQSKLSWRN GSWIPFYNKA
 LQYGDVIKLV GMAGEISDNF ALTSFKTKML AAHDTPIIAL NMGPAGKLSR VLNGFMTPVS HPDLPKAAAP
 QOLSATEIRQ ALALIGLEP KSFYLFKPI SSSRSPALHN TLFGATGLPH HYSLFETDKA SDVQDLIRAP
 HFGGASVTIP LKLDIVPLLD DISEAARAIG AVNTIIPIS AAKGDNKTTL LGDNTDWRGM VFALRSAGVA
 SPPGAGMVVG SGGTTRAAIY ALHSLGYAPI YVVARTPERV KELAGGF PAD YRIQH L ATRE EAVAVAVAAR
 HELPTVVIST IPADKPIDSG MREVLVTALQ GGAAGKEKG EPRVLEEMAY TPRHTPLMQL AEDAGWVTIP
 GLEVLAAQGW YQFQLWTGIT PLYADARA AV MGNESLERE RLEHHHHHH

Cloned by: Reinhard Albrecht

Curriculum Vitae

Harshul Arora

Department of Protein Evolution
Max Planck Institute for Developmental Biology
Max Planck Ring-5, Tuebingen 72076, Germany



Email: harshul.arora@tuebingen.mpg.de Phone: +49-1796001937

Born September 14, 1990, in New Delhi, India

Max Planck Institute for Developmental Biology Dec 2013- till date

Pursuing Ph.D., Department of Protein Evolution
Thesis: Multifunctional enzymes in the Shikimate Pathway
Supervisor: Dr. Marcus Hartmann

Madurai Kamaraj University 2010-2012

M.Sc. Biochemical Technology
Thesis: Genetic analyses of *slm* lesions in *Escherichia coli*: Role of *ssrA* and *arfA* in "SIR" phenotype
Supervisor: Prof. Dr. Hussain Munavar

University of Delhi 2007-2010

B.Sc. (H) Biochemistry Deshabandhu College

