STRUCTURAL INSIGHTS AND MOLECULAR DOCKING STUDIES FOR POLYCYCLIC AROMATIC HYDROCARBON DEGRADATION IN NOVEL BACTERIAL STRAINS FOR BIOREMEDIATION: AN IN-SILICO MODELLING STUDY

A thesis

Submitted in partial fulfilment for the award of the degree of

Master of Technology

In

Environmental Biotechnology

By

Ipsita Chakraborty

Reg No:154522

Exam roll No: M4EBT22001

School of Environmental Studies

Faculty of Interdisciplinary Studies, Law & Management (F.I.S.L.M)

Jadavpur University

Jadavpur, Kolkata

West Bengal 700032

June 2022

Declaration by student

My name is Ipsita Chakraborty, and I hereby certify that the work presented here is an original work and has not been published or submitted elsewhere to meet the requirements for a degree program. Any literature or work done by others has been acknowledged and well cited in the thesis in the reference section.

Ipsita Chakraborty

Roll No: 002030904001
Reg no: 154522
Exam roll no: M4EBT22001

Place: School of Environmental Studies, Jadavpur University

Date:

TO WHOM IT MAY CONCERN

It is hereby recommended that this thesis titled **"STRUCTURAL INSIGHTS AND MOLECULAR DOCKING STUDIES FOR POLYCYCLIC AROMATIC HYDROCARBON IN NOVEL BACTERIAL STRAINS FOR BIOREMEDIATION: AN IN-SLICO MODELLING STUDIES"** is prepared and submitted for the partial fulfillment of the continuous assessment of **Master of Technology in Environmental Biotechnology** course of Jadavpur university by **Ipsita Chakraborty(reg no: 154522, exam roll no: M4EBT22001)**, a student of the said course for the session 2020 – 2022, under my supervision and guidance. It is also declared that no part of this thesis has been presented or published elsewhere.

(Signature with Seal)
Director
Dr. Joydeep Mukherjee
School of Environmental Studies
Jadavpur University, Kolkata – 700032

(Signature with Seal)
Thesis Supervisor
Dr. Joydeep Mukherjee
School of Environmental Studies
Jadavpur University, Kolkata - 700032

(Signature with Seal)
DEAN
Faculty of Interdisciplinary Studies (FISLM)
Jadavpur University, Kolkata - 700032

ACKNOWLEDGEMENTS:

I would like to thank my guide Dr. Joydeep Mukherjee, Professor and Director of the School of Environmental Studies (F.I.S.L.M), for allowing me to conduct this study. The constant guidance he provided throughout the research and his constant suggestions are very much appreciated.

In addition, I am grateful for my co-guide, Dr Sayak Bhattacharya, to whom I owe a huge debt of gratitude for taking time out of his busy schedule to give me the logistical support I needed and for giving me valuable advice as I did my research.

I am grateful to my parents for their encouragement and support throughout my studies.

Finally, I extend my heartfelt thanks to all who supported me in my work directly or indirectly.

LIST OF ABBREVIATIONS:

PAH	Polycyclic Aromatic Hydrocarbon
US-EPA	United States Environmental Protection Act
DFT	Density-functional theory
RHD	Ring-hydroxylating dioxygenase
2D	2-dimensional
3D	3-dimensional
CO2	Carbon dioxide
H2O	Hydrogen dioxide
CH4	Methane
TCA	Tricarboxylic cycle
UM-BBD	University of Minnesota Biocatalysis/Biodegradation Database
DNA	Deoxyribonucleic acid
SMILES	Simplified molecular input line entry system
CESAR	Coding exon structure aware realigner 2.0
RNA	Ribonucleic acid
NCBI	National Center for Biotechnology Information
FASTA	Fast-all
PDB	Protein data bank
SOPMA	Self-optimized prediction method with alignment
PHYRE2	Protein homology/analogy y recognition engine
SNP	Single nucleotide polymorphism
SAVESv6.0	Structure validation server
HHS	U.S. Department of Health and Human Services
GRAVY	Grand Average of Hydropathy
II	Instability index
AI	Aliphatic index
MW	Molecular weight
SDF	Spatial data file
UFF	Universal force field
CSV	Comma-separated values
BLAST	Basic local alignment search tool

LIST OF TABLES:

LISTS OF FIGURES

TABLE OF CONTENTS

ABSTRACT

There has been an unprecedented rise in global pollution that has released many xenobiotic compounds into the environment. The threat of polycyclic aromatic hydrocarbons (PAH) as carcinogens and toxic pollutants is a major issue across the globe. Recent advancements in the application of bioinformatics in the field of bioremediation have paved the way in analyzing microbial metabolic ability and genetic insights for better PAH-degradation. Several bioinformatics tools have been developed over the last few years to assist in bioremediation procedures however the technique is still in its infancy. Not all PAH-degrading enzymes have been well-characterized, andhow they function within complex chemical and biological environments has not been extensively explored. The first part of this thesis discusses recent research findings on bacterial isolates, catabolic genes, enzyme degradation, and bioinformatics applications to bioremediation.

In the second part of the thesis, we used an in-silico approach to investigate and model 20 previously unknown degrading proteins from 10 novel bacterial species and assess their ability to bind to possible polycyclic aromatic hydrocarbon molecules. Further, we hypothesized that enzymes produced by bacterial species can degrade different PAH molecules equally, regardless of whether they have been shown to degrade experimentally. According to our findings, *Arthrobacter phenanthrenivorans Sphe 3* degrades not only phenanthrene but also napthalene and pyrene with highest efficiency. *Delftia acidovorans CS1-4* has the highest efficiency of degradation for phenanthrene, and *Alteromonas sp. SN2* has the highest efficiency for anthracene. All organisms show the same ability to degrade naphthalene except *Citreicella aestuarii 357* with lowest effiiency. Even though these degradation abilities haven't been established by experimental means, in-silico techniques have shown that they degrade. However, similar pattern of binding molecules was not observed in any protein-ligand complexes.

CHAPTER 1: INTRODUCTION

Biodegradation and bioinformatics are two interdisciplinary approaches to bioremediation. While bioremediation involves utilizing natural biological activity to aid in cleaning up contaminated sites [2], bioinformatics has the potential to accelerate environmental research, concerning global issues [3]. An unprecedented rise in global pollution has caused an increase in xenobiotic compounds in the environment [4].The threat of polycyclic aromatic hydrocarbons as carcinogens and toxic pollutants is a major concern worldwide. PAHs are benzene ring hydrocarbons that are produced by anthropogenic activity, 16 of which are classified as priority PAH compounds by US- EPA [5, 2]. Recent technology advancements have made these compounds much easier to detect, and their bioremediation procedure is now much more feasible due to the correct application of various bioinformatics tools and methods, overcoming the limitation of theconventional way of experimentation [7].

Several subfields of bioinformatics, as well as bioinformatics tools, are employed to identify degradation pathways and potential microorganisms for bioremediation [3]. Several microbes were identified, catabolic genes sequenced and their enzymatic reaction mechanism analyzed to assess their degradation efficiency [2,7]. Furthermore, many computational methods have been employed to study enzymes involved in degradation, including molecular dynamics, docking, DFT, and data mining [9]. Studies of the structure of the cytochrome P450 monooxygenase in microalgae such as *Haematococcus pluvialis* and *Parachlorella kessleri* have given insight into the potential role of protein conformation in enzymatic degradation [13]. The docking interaction between ligninolytic enzymes from basidiomycetes and PAH compounds has been extensively studied [24]. Computational methods have been extensively used to study ring-hydroxylating enzymes from *Spingomonas CHY-1* and their interaction with PAH [11]. Phylogenetic analysis of several algals, and halophiles has also been established in recent studies [12].

The use of bioinformatics in bioremediation is a new technique but it is still in its infancy [66]. Unfortunately, not all PAH-degrading enzymes have been well-characterized, and how they function within complex chemical and biological environments has not been

studied extensively. Thus, a deeper understanding of their molecular dynamics becomes increasingly crucial. A detailed analysis of the molecular characteristics of various proteins can explain their common interaction mechanism and how it differs from pollutants.

In the first part of this thesis, we review recent research findings on bacterial isolates, catabolic genes, enzyme degradation, and bioinformatics applications to the bioremediation of bacteria. The second part of this article examines 20 non-characterized degrading proteins originating from ten different organisms that degrade polycyclic aromatic hydrocarbons. Through the use of in-silico analysis, these enzymes were well-characterized and their degradation potential was analyzed by docking. Consequently, we hypothesized that bacterial species producing enzymes such as RHD and naphthalene dioxygenase that degrade different PAH molecules in different organisms; such organisms encoding these enzymes will also be able to degrade other forms of PAH equally, regardless of whether they have been shown to degrade experimentally. This facilitates the analysis of each microbe's degradation efficiency via the in-silico approach, overcoming the limitations of conventional methods. In the future, a genetically modified organism can be designed to encode a novel enzyme or protein that can degrade all 14 PAH pollutants more efficiently than current methods

CHAPTER 2: REVIEW OF LITERATURE

2.1. Chemistry and toxicity of PAH:

Polycyclic aromatic hydrocarbon is a carcinogenic and toxic organic pollutant. These are aromatic hydrocarbons containing 2 or more fused benzene rings, formed during incomplete combustion of fuels, during industrial processes requiring high temperatures, or by the addition of organic compounds at a temperature below 300°C [67]. Primarily produced by anthropogenic activities, they tend to be recalcitrant and have a toxic nature which increases with an increase in molecular weight [67]. PAHs are typically colorless, non-reactive, have a high boiling point, and have very low solubility in water. A growing number of fused rings decreases their solubility. These chemicals can volatilize and bio accumulates, causing serious health effects [2] Low-molecular-weight PAHs (such as anthracene, naphthalene, etc.) is associated with the vapor phase, whereas PAH with high molecular weight (such as pyrene) is difficult to degrade and considered genotoxic. There are hundreds of PAH compounds identified, outof which 16 are considered high-priority pollutants by the US-EPA [2]. The list of these priority pollutants is present in [figure 1]. How long PAH persists in the environment dependson multiple factors like physical and chemical properties, oxygen, concentration, bioavailability, etc. hence, before choosing specific bioremediation techniques, it is highly important to consider the limiting factors that affect PAH degradation [68].

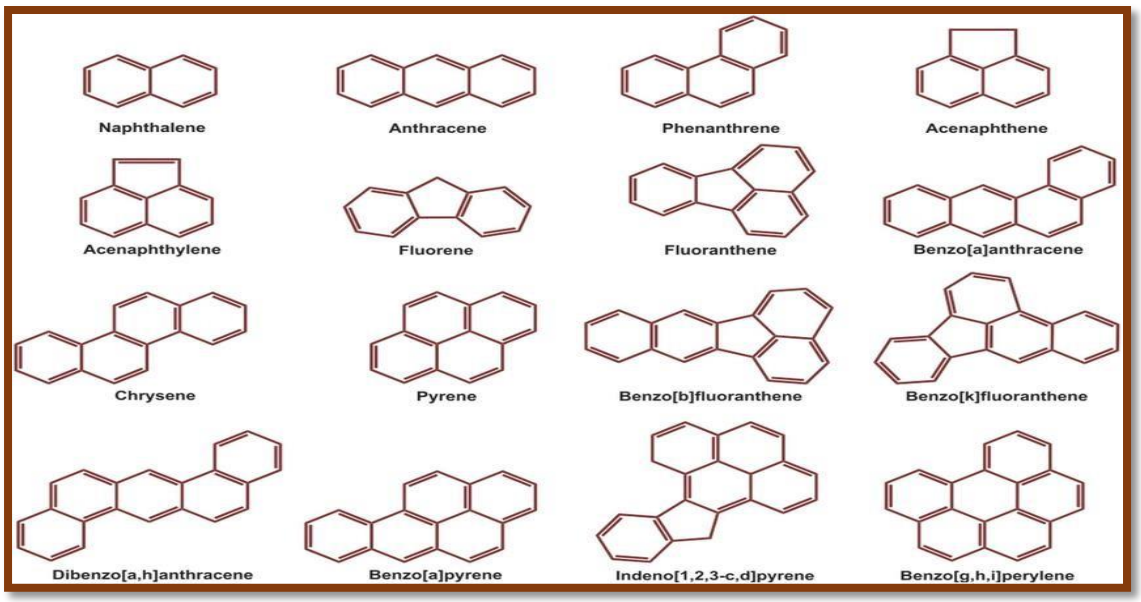

Figure 1: 16 priority PAH pollutants enlisted by US-EPA

The toxic properties of PAH compounds are due to their reactive metabolites. However, not all PAH compounds are toxic at the same level. Increasing the number of fused benzene rings in a compound increases its toxicity. (Cerniglia et al.,1992). The effective dose of the compound that interacts with the target site is another factor affecting toxicity (ATSDR 1995,2009) A substance's toxic response involves three levels: a response from the organism in form of death, abnormalities, edema, or abnormal growth; a mechanism of toxic action that involves teratogenicity, immunotoxicity, mutagenicity/genotoxicity, or carcinogenicity; and lastly, a mechanism of toxic action. Humans can suffer serious health effects from prolonged exposure to PAHs. Short-term exposure causes, irritation of the skin and eyes, nausea, and vomiting, whereas long-term exposure can lead to cancers of the breast, colon, prostate, kidney, liver, and cells [2].

Numerous studies have been conducted to investigate its mutagenic and carcinogenic properties (figure 2). For example, B(a)P is a highly toxic aromatic compound whose mutagenic, carcinogenic, and teratogenic properties make it one of the most dangerous substances ever discovered. According to Kristensen et al., 1995, pregnant mice exposed to high level of B(a)P gave birth to babies with birth defects and low birth weights. They are detected in marine organisms which can indirectly affect humans through food consumption. However, only a piece of minimal information about B(a)P degradation by microbes is available [68]. Because they have a low Ko/W value, they are adsorbed onto sediments and eaten by aqueous organisms [68]. They bio-accumulate in phytoplankton. They are readily absorbed through the skin and then into the human body, highlighting their bio-magnification in the food chain. According to Vanrooji et al., around 75 percent of total PAH enters our bodies through our skin because of its rapid absorption [68]. Upon entry into our body, PAH bind to cytochrome P450 enzymes in the liver and is converted into epoxides. These epoxides bind to DNA molecules causing mutagenic effects in the human body (figure 2) [2].

According to Sims and Overcash's (1983) findings, 1,2,3 ring structures have higher carcinogenic potential than other PAH compounds. Compounds such as Chrysene, diben(ah)anthracene, benzo(a)pyrene, benzo(a)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene and indol(1,2,3-cd)pyrene are classified as probable human carcinogens by the USEPA (2008). [2]. There has been an evaluation of PAH birth defects in African-American women from Washington Heights, Central Harlem, and South Bronx,

New York, and it has been found that exposure to PAH compounds during pregnancy leads to smaller heads and lower birth weights. (Perera et al.,2003) [2].

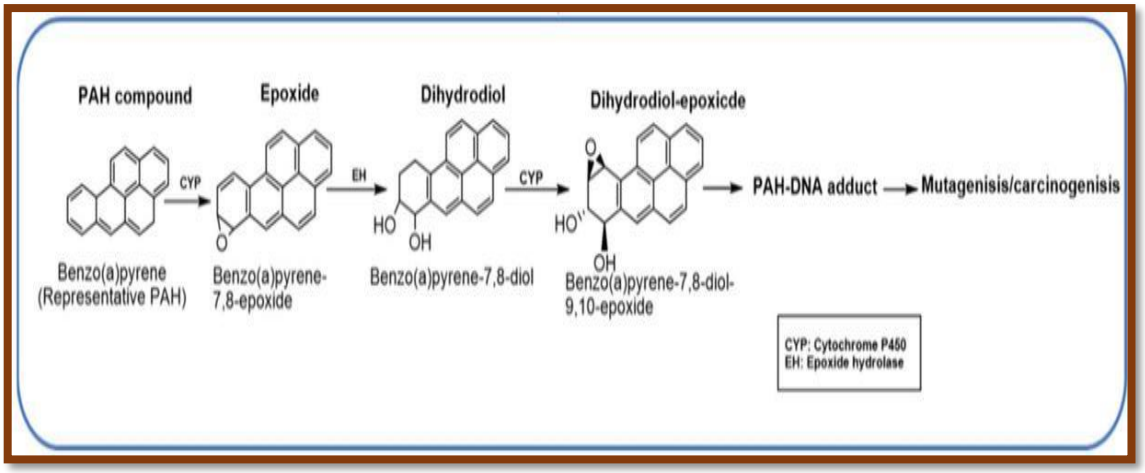

Figure 2: mechanism of PAH inside the human body

2.2. Application of bioinformatics in bioremediation of PAH:

Traditional methods such as chemical degradation, removal, adsorption, volatilization, and others seem to be simpler and more efficient. However, they have adverse impacts on the economy, human health, and biodiversity [69]. Recent years have seen a surge in popularity of biological treatment methods by bioremediation to degrade Polycyclic aromatic hydrocarbon.

Bioremediation is defined *as* an advanced technology by which microorganisms can biodegrade toxic or xenobiotic substances into a less toxic or non-toxic form. Bioremediation can be applied in two possible ways: 1) by enhancing the growth capacity of PAH-degrading microbes in the contamination zones *in situ*. 2) the addition of genetically modified organisms to degrade PAHs. However, the latter is rarely used [3].Algae,bacteria and fungi can degrade PAH compounds.They can bio transform complex PAH molecules into less complex compounds and then convert them into inorganic substances like CO_2, H_2O, and CH_4.

Effective PAH-degradation is mostly studied and applied to bacteria under appropriate conditions due to their diverse nature and easy adaptations. Microbes that are mostly studied for bioremediation are *Aeromonas, Enterobacter,Acinetobacter*

Corynebacterium, Bacillus, Pseudomonas, Sphingomonas, Micrococcus,Xanthomonas, Paenibacillus, Mycobacteria,. Microbes can degrade pollutants due to their genetic constituents and their associated proteins [2]. Bacteria degrade PAH compounds in a variety of ways - aerobic, anaerobic, and sometimes co-metabolizing (figure 3). Polycyclic aromatic hydrocarbons undergo successive steps of carboxylation, hydroxylation, and methylation where they are primarily activated before entering into their respective degradation pathway [68]. Aerobic degradation is accomplished by aerobic microbes and dioxygenase enzymes that break down compounds into metabolites that enter the TCA cycle to form co2, however, anaerobic degradation is accomplished by anaerobic bacteria that break down compounds under reducing conditions to metabolites that are oxidized further to form co2. The co-metabolism pathway occurs when a mixture of different PAH compounds is present in the contaminated sample. It involves the oxidation of non-growth substrates along with the growth of the microbes on a carbon source [68]. Using various bioinformatics tools like pathway prediction tools (PathPred and UM-BBD) helps in detecting these pathways and enzymes associated with them. Understanding the metabolic degradation pathways at the molecular level aids in formulating proper bioremediation [9].

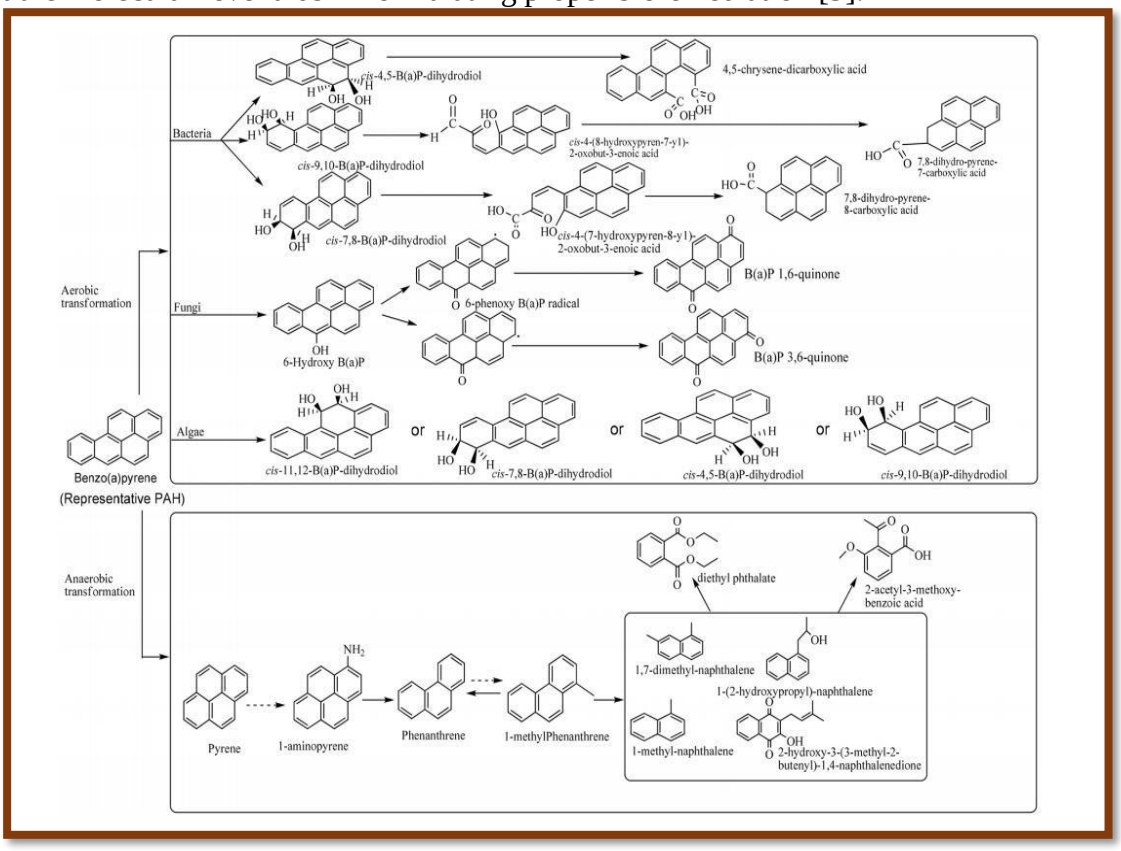

Figure 3: Aerobic and Anaerobic degradation of B(a)P

Hence, the identification of bacteria, the role of genes, and associated enzymes are important to be determined any bioremediation process. However, conventional experiments to be used to determine such parameters are very difficult for researchers with a very little level of accuracy. Also, analyzing these genes and proteins at the molecular level and under different microbial conditions is almost impossible to carry out for research purposes. Hence, bioinformatics comes into play. **Bioinformatics** combines biology with computational methods for dealing with biological systems at the cellular level. By doing this, one can look into the genomics and molecular levels of the biological system, overcoming the limitations of the conventional way of conducting experiments. [3]

Genomics, proteomics, computational biology, and system biology work together to assist in understanding the genetic basis of degradation and identifying novel degradation enzymes. Genomic research helps with understanding a gene's structure and function in an organism, based on the DNA sequence of the particular organism, whereas proteomics studies help in understanding the molecular mechanisms of bioremediation by analyzing the proteins in microbial cells. A complete genome study helps to understand the genetic basis, the production of secondary metabolites, and other physiological conditions. [2]. Computational biology integrates genomics and proteomics. It involves bioinformatics, which uses algorithms and statistical techniques to interpret, classify, predict gene expression, and build models of molecules of biological significance. System biology models interactions between proteins on a large scale. It also enables establishing the phylogenetic relationship between organisms that are important concerning the bioremediation perspective. These sub-fields of bioinformatics assist in understanding reactivity patterns and other parameters associated with degradation [3].

Bioinformatics when used in bioremediation studies aids in the following ways: [8]

- Understanding physical and chemical features of degradation proteins and compounds. through several databases that provide information such as SMILES code, chemical structure in PDB format, canonical structures, formulas, etc.
- Determining toxicity level of compounds through toxicity determining databases like CAESAR, GENE-TOX, ECO-TOX, etc.

- Determining the degradation pathway of microbes via pathway prediction tools.

- Microorganisms, catabolic genes, and possible enzymes involved in pollutant degradation are available in various databases.
- Developing protein structures and models based on similarity studies via several databases.
- Establishing phylogenetic studies through various software like MEGA etc.
- Understand molecular interaction mechanism and dynamics through in-silico methods like protein-ligand docking etc.
- Using several bioinformatics tools for primer designing, and restriction fragments for isolating degrading microorganisms.

Several bioinformatics tools have been developed over the last few years to has assisted in bioremediation procedures. These bioinformatics tools are further classified based on the functions they perform, details of which are mentioned in figure 4. These tools not only help in identification but also in performing in-silico characterization of bacteria and proteins involved and degradation efficiency. Some of these tools are used in this research process[3].

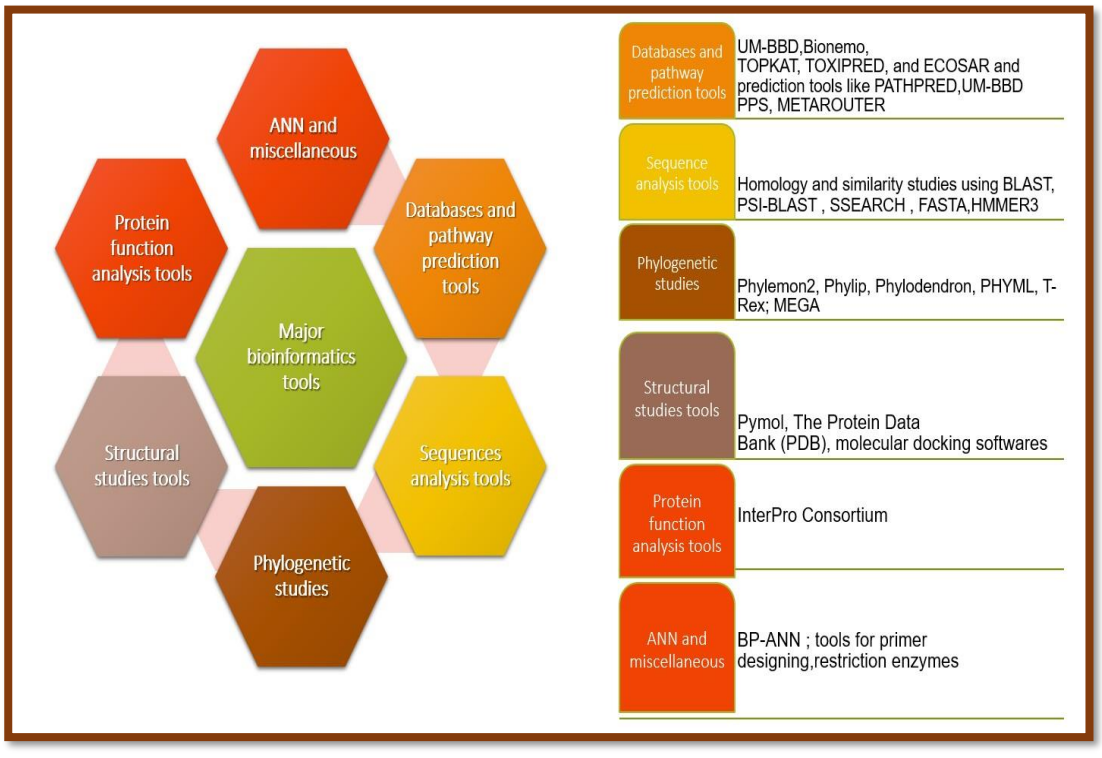

Figure 4: bioinformatics tools applied to bioremediation of Polycyclic aromatic hydrocarbon

2.3. Enzymes involved in biodegradation:

Enzymatic degradation of PAH is much more efficient than other methods of degradation, such as chemical degradation, due to its stereospecificity, mild reaction conditions, and high reaction rate. The role of PAH-degrading enzymes has been the subject of numerous studies. Enzymes like mono- and di-oxygenases, manganese peroxidases, dehydrogenases, and laccases are reported. Dioxygenases usually contain several subunits such as ferredoxin, reductase, and terminal oxygenase[2]. A cytochrome P-450 monooxygenase is an oxidoreductase that can degrade PAHs via a cytochrome P-450-mediated pathway. Among the known microbial P450s, CYP1A1, CYP102, CYP1A2, CYP101, and CYP1B1 can metabolize PAHs. In studies of CYP102A1 from Bacillus megaterium, (P450BM3), it was demonstrated that it could oxidize fluoranthene,phenanthrene, pyrene to phenols and quinones [70]. Kan et al., isolated Rhodococcus sp. P14 from crude oil-contaminated sediments whose expression of CYP108J1 resulted in PAH degradation, that uses PAH as a carbon and energy source[70]. Kan et al., isolated Rhodococcus sp. P14 derived from crude oil-contaminated sediments, whose expression of CYP108J1 led to PAH degradation, which utilizes the PAH as a carbon and energy source [70].

Ring hydroxylating dioxygenases (RHD) catalyze the first oxidation step resulting in dihydriol metabolitesThe ring-hydroxylating dioxygenase phn1 in Sphingomonas CHY-1 can oxidize PAHs with four and five fused rings, as well as high molecular weight PAHs [66,70]. Diverse forms of RHDs such as phthalate dioxygenase, naphthalene dioxygenase, biphenyl dioxygenase, and catechol dioxygenase have been extensively characterized and studied [2]. In *Nocardioides sp*. Gene cluster phdEFABGHCD encodes phenanthrene dioxygenase was investigated. A plasmid- carrying gene ABCD was introduced into E. coli and the transformant carrying this gene is capable of degrading phenanthrene compounds. Genetic analysis of this gene cluster in Nocardioides sp. Identified a novel RHD enzyme called phenanthrene dioxygenase that helped in phenanthrene degradation[2].

Aldehyde dehydrogenase appears to be involved in the metabolism of aromatic compounds[70]. The aldehyde dehydrogenase (Nidd) catalyzes the degradation of 1-hydroxy-2-naphthaldehyde to 1-hydroxy-2-naphthaloic acid [70]. A study of the protein

expressed in *Amycolatopsis tucumanensis* DSM 45259 showed that aldehyde dehydrogenase was abundantly expressed [70]. The denitrifying bacterium

Azoarcus sp. produces a specific enzyme called EBDH, which is enantioselective. Aromatoleum aromaticum (EbN1) is capable of catalyzing the oxygen-independent, stereospecific hydroxylation of ethylbenzene to 1-(S)-phenylethanol; this is the direct anaerobic oxidation of a non-activated hydrocarbon and is crucial to biomineralization.Several such investigations have been conducted, leading to the discovery of several novel enzymes based on ring oxidation and cleavage products [70]

2.1. Catabolic Genes involved in biodegradation:

Information about the catabolic genes is critical to understand its metabolism but also about the evolution of diverse degradation pathways and the structural and functional relationship of catabolic genes. 16S rRNA sequencing helps to identify microbes from contaminated zones. Sequence alignments and phylogenetic analyses help to identify species with similar genetic makeup and determine their evolutionary relationship[2]. The majority of PAH degradation-related genes are common to several aromatic hydrocarbons, and these genes code for degradation-related enzymes. Polymerase chain reaction and DNA hybrid analysis help to detect those catabolic genes that assist in degradation[2]. These catabolic genes categorized as gene clusters (Kiyohara et al.,) code for enzymes that play a pivotal role in PAH-degradation. There are diverse forms of PAH, each of which is degraded by various degrading enzymes coded by these catabolic genes. Different catabolic clusters like *pah, Nah, nid, nag, dox, phn, nar, pdo, etc.*, were reported from *Pseudomonas, Burkholderia, Sphingomonas, Mycobacterium*, etc. [2]. Usually, these genes are present on chromosomes or plasmids.

The nahAc gene is conserved vastly in many gram-negative bacteria due to horizontal gene transfer among diverse bacterial species (Stuart-Keil et al.1998). In Pseudomonas putida G7, naphthalene catabolic genes have been well characterized (Simon et al., 1993.In a plasmid NAH7, naphthalene catabolic genes(nah) are divided into two operons that encode enzymes required for the lower and upper pathways of naphthalene cycle. In the lower pathway, these enzymes convert salicylate into pyruvate and acetaldehyde (Simon et al.,1993)[71]. A common transcriptional regulator, NahR, a

type of LysR, positively regulates the operons In bacteria,salicylate increases the expression of NahR, resulting in higher levels of nah gene expression(Yen and Gunsalus, 1985; Peng et al., 2008). in a well-studied analysis of low molecular weight PAH by Simon et al., on pseudomonad species, nah-like genes for naphthalene and phenanthrene are highly conserved and are 90 percent similar to nah genes of pseudomonad putida G7[71].

In Ratstonia sp. U2, the naphthalene catabolic operon (nag) contains all genes of the upper pathway in a similar order to those found in Pseudomonas strains, except two additional genes named nagG and nagH, which are structural subunits of thesalicylate-5-hydroxylase enzyme that converts naphthalene to gentisate (Zhou et al., 2001). Comamonas testeroni strain GZ42 also possesses genes for the catabolism of naphthalene, similar to those found in Ralstonia sp. U2 (Goyal and Zylstra, 1997).[66] There are similar genes involved in the lower pathway of naphthalene catabolism in several Pseudomonas strains (NCIB9816-4, ND6, P. putida G7, and P. stutzeri AN10 (Hake and Omori, 2003; Peng et al., 2008). Among the 11 genes in the naphthalene operon, nahY represents the naphthalene chemotaxis gene[66]. AN10 and ND6 strains, however, contain a second salicylate hydroxylase gene (nahW), which is outside the salt operon. The pah catabolic genes of spingomonads show high levels of structural and functional similarity[66].

A gene cluster encoding for ring hydroxylating dioxygenase enzyme in Rhodococcus sp. reported that its expression level increased with the addition of mixtures of PAH compounds such as pyrene, anthracene, phenanthrene, etc. (Peng et al. 2018)[66,2].

Besides all the other genes that have been used for degradation, we have selected two others that encode degrading proteins that are promising in bioremediation procedures and considered suitable biomarker. These genes are pahAC and pahE[1]. To understand the efficiency of the microbes, it is important to identify their native degrading enzymes responsible for biodegradation. PahAC, which encodes the alpha subunit of PAH-ring hydroxylating dioxygenase, has been used as a functional marker gene for PAH-degrading bacteria in environmental samples. however, Poor phylogenetic resolution and a lack of specificity for PAH resulted in an underestimation for degrading bacteria. Hence, another superior biomarker pahE gene that codes for the hydratase-aldolase enzyme are also selected as the enzyme of this gene catalyzes the 5th step of aerobic degradation by converting trans-o-hydroxybenzylidenepyruvate into aldehyde and

pyruvic acid.

Among the PAH- degrading bacteria, pahE gene was most effective as a functional marker. Both these genes have been been been shown to identify degrading bacterial consortium. However, till now very few studies have been conducted on the proteins these genes code in different organisms and thus allows us to explore them at a molecular level[1].

In Literature 1, both *pahAC* and *pahE* genes as molecular markers have been thoroughly studied. There have been around 73 microbial organisms studied, 68 of which were specifically targeted, and phylogenetic relationships have been established. While these microorganisms and their associated proteins have been identified, not all have been well characterized. Some have simply been sequenced, while others have degraded PAHs experimentally, but their degradation efficiency has not been evaluated yet. There has been a thorough literature review of these organisms. 47 organisms were identified that are capable of degrading PAH priority pollutants out of the total 73 identified. For our research project, ten of these novel organisms are selected, whose enzymes are characterized and the efficiency of their degradation is predicted.

The use of bioinformatics in bioremediation is a new technique for studying its mineralization process, but it is still in its infancy. Limited studies are focusing on the biodegradation of priority PAH pollutants especially high molecular weight PAHs in terms of PAH bioremediation. Only a few microbes have been identified that are capable of degrading. In addition, the degradation efficiency of microbes found so far has not been evaluated, which is crucial for developing any bioremediation procedures. Researchers find x-ray crystallography and NMR technology for protein structure costly and inconvenient when it comes to evaluating degradation efficiency. However, such limitations can be removed by proper application of bioinformatics which is not only cheap and convenient but also provides elaborate understanding for developing any hypothesis.

Table 1:Microorganisms, catabolic genes, and enzymes involved in bioremediation

PAH-catabolic gene	Microbial strain	Target PAH	Reference
Nah	*Pseudomonas putida* strains G7	Naphthalene	Simon et al. (1993)
	Sphingomonas yanoikuyae B1	Naphthalene, phenanthrene	Zylstra and Kim (1997)
	Lolium multiflorum L.	Pyrene	Khan et al. (2009)
	Pseudomonas putida KT2440	Naphthalene	Fernández et al. (2012)
	Sphingomonas koreensis Strain ASU-06	Naphthalene, phenanthrene, anthracene, pyrene	Hesham et al. (2014)
pah	*Pseudomonas putida* OUS82	Naphthalene, phenanthrene	Takizawa et al. (1994), Kiyohara et al. (1994)
	Pseudomonas aeruginosa PaK1	Naphthalene	Takizawa et al. (1999)
dox	*Pseudomonas* strain C18	Naphthalene	Denome et al. (1993)
	Arabidopsis thaliana AT5G05600	Phenanthrene	Hernández-Vega et al. 2017
Phn	*Burkholderia RP007*	Naphthalene, phenanthrene	Laurie and Lloyd-Jones (1999a, b)
	Acidovorax strain NA3	Phenanthrene, naphthalene, chrysene, benz[a]anthracene, benzo [a]pyrene,	Singleton et al. (2009)
	Burkholderia sp. Ch1-1	Phenanthrene	Hickey et al. (2012)
	Delftia acidovorans Cs1-4	Phenanthrene	Shetty et al. (2015)
phd	*Nocardioides* sp. KP7	Phenanthrene	Saito et al. (1999, 2000)
	Nocardioides sp. KP7	Phenanthrene, 1-methoxynaphthalene	Chun et al. (2001)
	Mycobacterium vanbaalenii PYR-1	Pyrene	Kim et al. (2007)
nag	*Ralstonia*sp. strain U2 (formerly *Pseudomonas* sp. strain U2)	Naphthalene	Fuenmayor et al. (1998), Zhou et al. (2001)
	Polaromonas naphthalenivorans CJ2	Naphthalene	Jeon et al. (2006)
	Herbaspirillum sp. strain RV1423	Naphthalene	Jauregui et al. (2014)
nid	*Mycobacterium* sp. Strain PYR-1	Anthracene, fluoranthene, pyrene, phenanthrene, and benzo[*a*]pyrene	Khan et al. (2001)
	Mycobacterium vanbaalenii PYR-1	Naphthalene, phenanthrene, anthracene, fluoranthene, pyrene, benz[a]anthracene	Kim et al. (2006)
	Mycobacterium sp. strain CH-2	Phenanthrene, pyrene, and fluoranthene	Churchill et al. (2008)
nar	*Rhodococcus* sp. NCIMB12038	Naphthalene	Larkin et al. (1999)
	Rhodococcus sp. NCIMB12038	Naphthalene	Kulakov et al. (2000)
pdo	*Mycobacterium* sp. strain 6PY1	Pyrene	Krivobok et al. (2003)
	Mycobacterium sp. strain CH-2	Phenanthrene, pyrene, and fluoranthene	Churchill et al. (2008)
	Lolium multiflorum L.	Pyrene	Khan et al. (2009)
C12O	*Pseudomonas* sp. CZ2	Naphthalene	Zhou et al. (2013a, b)
	Sphingomonas koreensis Strain ASU-06	Naphthalene, phenanthrene, anthracene, pyrene	Hesham et al. (2014)
	Celeribacter indicus P73T	Fluoranthene	Cao et al. (2015)
	Pseudomonas aeruginosa, *Pseudomonas* sp., *Ralstonia* sp.	Phenanthrene, pyrene, fluoranthene	Sangkharak et al. (2020)
C23O	*P. putida* OUS82, *P. putida* NCIB 9816, *P. stutzeri* AN11, *Sphingomonas yanoikuyae* DSM 6900, *P. puorescens*, *Sphingomonas* sp., *P. merzdocina*	Naphthalene and phenanthrene	Meyer et al. (1999)
	Pseudomonas sp. Strain ND6	Naphthalene	Jiang et al. (2004)
	Pseudomonas sp. CZ2 and CZ5	Naphthalene	Zhou et al. (2013a, b)
	Sphingomonas koreensis Strain ASU-06	Naphthalene, phenanthrene, anthracene, pyrene	Hesham et al. (2014)
	Pseudomonas aeruginosa, *Pseudomonas* sp., *Ralstonia* sp.	Phenanthrene, pyrene, fluoranthene	Sangkharak et al. (2020)
fld	*Sphingomonas* sp. LB126	Fluorene	Wattiau et al. (2001)

(*Source: A.K Hartish et al,2020)*

CHAPTER 3: OBJECTIVES AND LIMITATIONS

3.1.Objectives:

A **major objective** of the study is to characterise 20 previously uncharacterized degrading proteins from novel bacterial organisms and to assess their ability to bind to their likely polycyclic aromatic hydrocarbon molecules via in-silico approach. The **specific objectives** of the study are:

a. Analyze the conformation of all degrading proteins in both two-dimensional and three-dimensional form from different bacterial strains.
b. Perform characterization studies on each of the 20 proteins.
c. Conduct molecular docking analyses to determine their binding potential to their respective ligands. Thus, analyse each microbe's degradation efficiency.
d. Construct hypotheses regarding the binding potential of enzymes to unknown ligands not shown to degrade experimentally
e. Visualize all docked molecules to identify pocket atoms and interacting amino acids involved to check for conserved residue

3.2. Limitations of the study:

a. Due to the fact that all structural conformations of proteins are determined by in-silico software and not by X-RAY diffraction and NMR methods, the structures cannot be included in structural databases.

b. Blind docking method has been used for docking process since active sites of the molecules were unknown. Active site determination is important to go for any in-depth binding studies, However, such method can be used in research in order to analyse the degrading potential of any enzyme.

c. Lastly, all the data obtained thorough in-silico approach gets further confirmed and validated by experimental techniques where real time environment is involved.

CHAPTER 4: MATERIALS AND METHODS

4.1 Datasets:

Ten bacteria of different genera were selected for our study based on their ability to degrade PAH compounds, as described in the literature [1]. We conducted a detailed literature review across different databases to find out the potential PAH compounds the bacteria could degrade experimentally. The primary amino acid sequences of all degrading proteins are obtained from NCBI's protein database [1] by referencing the accession numbers available in the literature [1] and are downloaded in FASTA format. The three-dimensional structures of the proteins under study are then searched in the PDB database. Although no equivalent tertiary structures were found in the structural database, the proteins must be considered novel and require structural and functional analysis.

4.2. Predicting Secondary Structure:

The secondary structure of 20 protein sequences as degrading enzymes is predicted using SOPMA software (https://npsa- prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) based on the amino acid sequences of the proteins. A protein's 2D structure refers to its polypeptide backbone. SOPMA predicts the secondary structure of proteins using a sequence similarity-based algorithm. The prediction accuracy is 69.5% in 3 conformational states (alpha helix, beta sheets, turns). Predictive parameters optimize the accuracy of the prediction [14]. The predictive parameters include several threshold states, similarity threshold and window width. Our proteins of interest are modelled with 4 threshold states, 8 similarity thresholds, and a 17-window width with an output width of 70.

4.3.Predicting Tertiary Structure:

Since no experimental 3 dimensional structures of all 20 proteins were available, the Protein Homology/Analogy Y Recognition Engine (PHYRE2) (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) server was used to predict and visualize their tertiary structures. Having a 3D structure of a protein provides much deeper insights into the function of the protein. Phyre2 uses remote homology detection techniques to build 3D models. predicting ligand binding sites, and determining the impact of amino acid variants (such as nsSNPs) on a protein sequence. Using a simple user interface, allows a user to submit a protein sequence interpret the 2D and 3D structures, compositions in domains, and model quality of their models. Keeping the modelling mode at NORMAL, all the 20 amino acid sequences of proteins are analysed to generate their hypothetical 3D structures[15].

4.4. Validating Modelled Proteins:

The hypothetical protein models generated are validated using validation software like PROCHECK [23], and ERRAT [20], all at once from the SAVES v6.0 server(https://saves.mbi.ucla.edu/) to check their structural quality and whether it fits for further analysis. PROCHECK analyses the stereochemical properties of proteins. The method examines how normal or unusual a protein structure's residue geometry in comparison with stereo-chemical parameters derived from detailed, high-resolution structures. [17]. The first plot that it produces is a Ramachandran plot, which represents the statistical distribution of phi-psi torsional angles of all residues in a protein. [17] Phi/Psi distributions can be used to verify structure validity in protein structures [16]. The plot separates regions into allowed and disallowed regions for dihedral angles, allowed being the most favourable phi-psi values. It is expected that high-quality models will have more than 90% in allowed regions [17]. ERRAT software evaluates the non-bonding interactions between different atoms in the modelled protein and plots the data by comparing it with statistical data from highly refined experimental structures, giving a score which represents the quality of the modelled protein [18, 19]. The accepted range of the score should be greater than 50 to be considered a

good model [19]. All the protein structures in the SAVES server are to be submitted in PDB format.

4.5. In-Silico Characterisation:

4.5.1. Presence Of Signal Peptide:

Signal peptides are short sequences of peptides found in the N-terminal region of a protein, which aid in protein targeting and protein translocation. Our protein molecules were predicted to contain signal peptides using SignalP4.1 (https://services.healthtech.dtu.dk/service.php?SignalP-4.1). [25] SignalP6.0(https://services.healthtech.dtu.dk/service.php?SignalP), which is an updated version of SignalP, not only predicts the presence of signal peptides but also mentions 5 types of signal peptides present and their cleavage sites based on protein language models [24]. In our study, we used both these servers for better analysis.

In the case of signalP4.1, the software asks for the FASTA sequence of the primary amino acid sequence of our proteins. The user must indicate what type of organism (gram-positive, gram-negative, Eukarya) secretes the protein. Keep the D-cut-off value as “DEFAULT” and save graphics as “PNG” s. Keep the output format as “STANDARD” for all proteins. It is best to select "Input sequences may include TM regions" for our search METHOD, and then SUBMIT it. However, for signal6.0, the output format should be “LONG” and the model mode should be “FAST”.

4.5.2. Domain Analysis:

Domain analysis of proteins is carried out using the PFAM database(https://pfam.xfam.org/search#tabview=tab1), which classifies proteins into families, domains, and clans in response to a large sequence collection based on UniProt Knowledgebase (UniProtKB) [26]. The relationship between entries is determined via comparisons based on similarity in sequence, structure, function, and profile-profile comparisons using software such as HHsearch [27,28]. Our sequences of the protein were uploaded as FASTA files to

batch sequence search mode. The cut-off E-value should be kept at 1.0 and provide our email address to receive our results.

4.5.3. Prediction Of Physio-Chemical Parameters:

EXPASY PROTPARAM server(https://web.expasy.org/protparam/) was used to explore the physiochemical properties of all enzymes according to Gasteiger et al., from a protein's amino acid sequence [29]. A list of physio-chemical parameters is provided by the software, The following parameters are described below [29]:

***A.* Extinction coefficients:** A protein's extinction coefficient indicates how much light it absorbs at a given wavelength. This coefficient is useful to know when a protein has been purified by a spectrophotometer. Calculated by the formula-

E.C of protein = No. of (Y) Ext (Y) + No. of (W)*Ext(W)+No.of (C)*Ext(C)*

Ext(Y) = 1490, Ext(W) = 5500, Ext(C) = 125 at absorbance of 280 nm

where absorbance is calculated by E.C of protein / Molecular weight

B. In vivo half-life: This is a measure of how long it takes for half of the protein synthesized in a cell to disappear. In ProtParam, the half-life of a protein is predicted using the N-end rule based on the N-terminal residue of three model organisms (human, yeast, and E. coli).

C. Instability index (II)

Instability indexes measure how stable your protein in a test tube is. According to a statistical analysis of 12 unstable and 32 stable proteins, certain dipeptides occur at significantly different rates in unstable and stable proteins. Instability index of a protein is calculated by -

i=L-1

position of dipeptide = L-1 where L= length of sequence.

*Instability index(II) = (10/L) * SUM DIWN(X[I]*[i+1])*

where DIWV(x[i]x[i+1]) = instability weight value for the dipeptide startin g in position i.

D. Aliphatic index: The aliphatic index of a protein refers to the volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). In terms of increased thermostability of globular proteins, it can be regarded as a positive factor. The aliphatic index of a protein can be calculated as follows:

*Aliphatic index (AI) = X(A) + a * X(V) + b * (X(I) + X(L))*

where X(A), X(V), X(I), and X(L) are mole percent (100 X mole fraction) of alanine, valine, isoleucine, and leucine.

a = relative volume of the valine side chain (a = 2.9)and

B= relative volume of Leucine and Isoleucine side chains (b = 3.9) to the side chain of alanine.

E. GRAVY (Grand Average of Hydropathy): The GRAVY value of a peptide or protein is determined by the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence.

Molecular weight and atomic composition are also provided by the software [30]. All of the software's parameters are calculated using statistical mathematical data, which indicates its high accuracy. For our analysis, the protein amino acid sequence is provided in FASTA format.

4.6. Molecular Docking Analysis And Visualization:

By using PYRX software(https://pyrx.sourceforge.io/home), docking analyses were performed to determine how different protein molecules bind with their degrading compounds. Following that, docked molecules are visualized in BIOVIA DISCOVERY STUDIO VISUALIZER [37] (https://discover.3ds.com/discovery-studio-visualizer-download). Both PYRX and DISCOVERY STUDIO software can be downloaded from the official website for free. As input files to the PYRX software, the protein molecules in our study are loaded in PDB format. To perform further screening, protein molecules must be converted to PDBQT files as macromolecules. Before docking, it is essential to download ligands from PubChem(https://pubchem.ncbi.nlm.nih.gov/).

In our study, we need the ligands naphthalene [32], phenanthrene [33], anthracene [34], pyrene [35], and pyruvate [36], whose 2D conformers can be downloaded as an SDF file. In the control panel, click the “OPEN BABBLE” option to load the desired ligand using the (+) sign(r). The ligand molecule in SDF format is added to the site where the ligand shall be energy minimized. Energy minimization parameters include force fields and optimization algorithms. The default value for the force field option is "UFF", which stands for the universal force field. As part of the “optimization algorithm”, we need to select “conjugate gradients” and their default values for each sub-parameter. the energy minimized ligands are then converted to PDFQT files as AUTODOCK ligands [31]. Click on the Start button under the Vina Wizard tab in the Controls panel. Select the ligands and molecules that need to be docked and click on "FORWARD"[31]. A grid box appears. Since we are doing blind docking, Click on "MAXIMIZE" to maximize the grid box or adjust the grid to fit the entire protein molecule. The analysis will begin. Docking results are saved as CSV files, and docked molecules are saved as PDB files.

The docked proteins are loaded into DISCOVERY STUDIO [37] as input files, where their interactions are visualized in both 2D and 3D. In this study, the interactions between the interacting atoms of the proteins with their ligand and the types of interaction between the proteins and their ligand were noted for analysis.

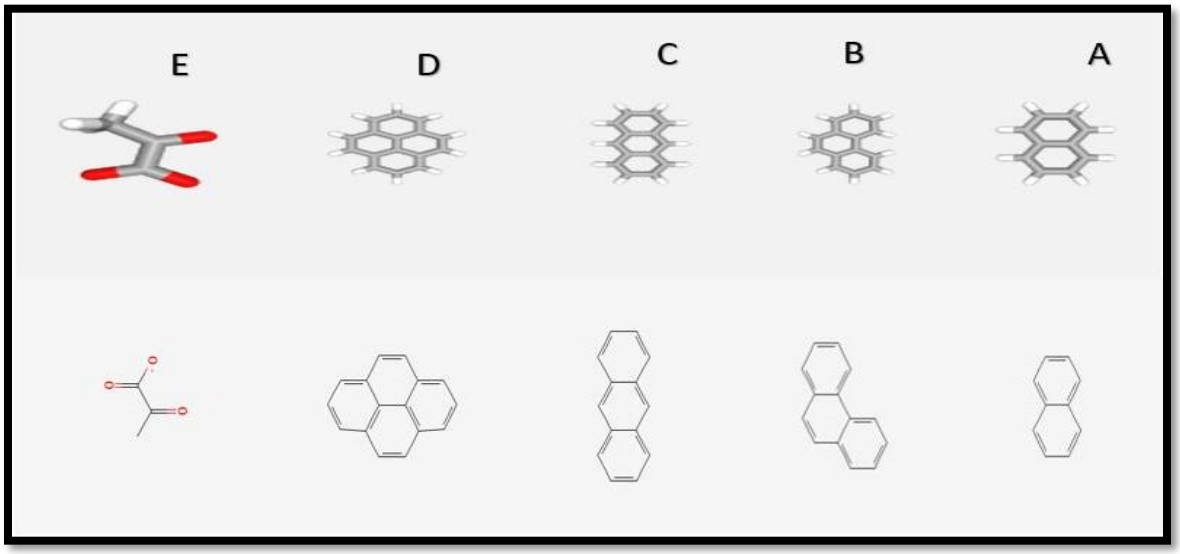

Figure 5: 2D (below) and 3D conformations (above) of ligands : A: naphthalene B: phenanthrene C: anthracene D: pyrene E: pyruvate

CHAPTER 5: RESULTS AND DISCUSSIONS

5.1. Bibliography of The Bacterial Species Under Study

Among the factors determining the purpose of our research is the analysis of microorganisms, especially bacteria, and their degrading enzymes, which have the capability of destroying PAHs. Literature (1) has already published a list of bacteria from various genera with the ability to degrade PAH, whose phylogenetic analysis has been explained extensively. An in-depth literature review was conducted on these bacteria and the enzymes (1) they encode. We have seen that only a few bacteria have been well-characterized both at a cellular and molecular level.

Specifically, we are interested in those other bacteria that have been experimentally shown to degrade, but whose proteins haven't been well characterized. Such 10 bacterial species selected for our study are- ***Alcaligenes faecalis AFK2, Alteromonas sp. SN2, Arthrobacter phenanthrenivorans Sphe 3, Burkholderia sartisoli rp007, Citreicella aestuarii 357, Delftia acidovorans CS1-4, Polaromonasnaphthalenivorans strain CJ2, polycyclovorans algicola TG408, Ralstonia sp. Strain U2, Terrabacter sp. Strain DBF63.*** These bacteria specifically target those polycyclic aromatic hydrocarbons that are considered priority pollutants by the US EPA. [Table 2].

It is critical to note that, unlike previous studies, our approach did not explore bacteria's biochemistry or genetic makeup. Specifically, this study examines the characterization of such proteins and how they break down compounds at the molecular level through in-silico approaches.

Table 2: lists of all organisms selected, the enzymes they encode, and possible PAH pollutants theydegrade

Organisms	Enzymes they secrete		Pah degraded	Literature
	pahAc	***pahE***		
Alcaligenes faecalis AFK,2	Dioxygenase large subunit	Hydratase-aldolase	Phenanthrene	[39] Kiyohara et al.,1982
Alteromonas sp. SN2	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Naphthalene, phenanthrene,anthracene, pyrene	[40] Hyun Mi-Jin et al.,2012
Arthrobacter phenanthrenivorans Sphe 3	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Phenanthrene	[41] Elpiniki et al.,2014
Burkholderia sartisoli rp007	Dioxygenase alpha subunit	Hydratase-aldolase	Naphthalene, phenanthrene and anthracene	[42] Vanlere et al.,2008
Citreicella aestuarii 357	IPB-dioxygenase, ISP large subunit	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	Napthalene	[43] Bosch et al., 2013
Delftia acidovorans CS1-4	Napthalene1,2-dioxygenase	Dihydrodipicholinate synthetase	Phenanthrene	[44] Shetty et al., 2015
Polaromonas naphthalenivorans strain CJ2.	Naphthalene	Aldolase	Naphthalene	[45] jeon et al., 2004
polycyclovorans algicola TG408	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Naphthalenee	[46] Thompson et al., 2018
Ralstonia sp. Strain U2	Napthalene1,2-dioxygenase	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	Naphthalenee	[47] Jones et al., 2003
Terrabacter sp. Strain DBF63	Phthalate dioxygenase	None	Fluorene	[48] Habe et al.,2004

5.2. Secondary, Tertiary Structures And Their Validation

5.2.1. Secondary Structures:

Proteins are mainly composed of three levels of structural organization. There are primary, secondary, and tertiary structures. The primary structures are amino acid sequences linked together with peptide bonds to form long polypeptide chains[49]. The secondary structure consists of regions that are stabilized by hydrogen bonds between partially negative oxygen atoms and partially positive nitrogen atoms in the polypeptide backbone. Proteins form their secondary structure by folding or coiling their polypeptide chains into conformations that contribute to their structural stability. Usually, these conformations are referred to as alpha-helixes, beta-sheets, coils, and turns [50]. Alpha helices are single chain, right-handed coils that are joined together by intramolecular hydrogen bonds between every fourth amino acid that gives it a spiral shape [50]. This conformation is the most common among all protein secondary structures. The alpha helical regions formed by it are identified by its distribution of φ and ψ dihedral angles in protein conformation. The alkyl groups of the alpha-helix are located outside the helix. In addition to alpha helices, proteins also contain β-sheets. β-sheets are prepared by stretching segments of a polypeptide chain and forming intermolecular hydrogen bonds between adjacent strands [51]. the stability of hydrogen bonds is much higher compared to the alpha helix. The alkyl groups are located both inside and outside of the beta strand [50]. The loop formed by two strands that change direction abruptly is called a β-turn. Turns are longer than loops and they connect other secondary structures. The loop regions of proteins are often highly variable. In sequence alignment and homology modeling when it is critical to have an accurate alignment of amino acids, loop regions offer a good candidate for detection of insertions and deletions in amino acid sequences. [50]. A protein's fold is closely related to the combinations of helices, helices and strands, etc. Consequently, identifying secondary structure elements and structural motifs is an advantage when viewing a 3D protein structure.

There are mostly alpha helices, beta strands, beta turns, and coils in the proteins under study, except for *Citreicella sp., Terrabacter sp., and Arthrobacter sp.* [see table 2]. *Arthrobacter sp.* has a ring-hydroxylating dioxygenase enzyme that has no turns in comparison to the other organisms. As well, it has the lowest affinity for phenanthrene compared with other organisms, which might indicate the absence of coils in its secondary

structure can affect its ligand binding. However, the proteins that bind with naphthalene in different organisms are either similar in their secondary conformations or have similar binding affinities with naphthalene regardless of the organism or protein they bind. There is no significant difference in protein's secondary structure depending on the type of ligand they bind to [see table 3]. Similarly is in the case of all pahE genes.[see table 4]

Table 3: secondary structure of all enzymes encoded by pahAc gene

Bacteria	***pahAc***	**helix %**	**extended strands%**	**coils%**	**turns %**	**others %**
Alcaligenes faecalis AFK2	dioxygenase large subunit	33.41	17.85	41.88	6.86	0
Burkholderia sartisoli rp007	dioxygenase alpha subunit	32.44	18.22	44.89	4.44	0
Citreicella aestuarii 357	IPB-dioxygenase, ISP large subunit	13.3	25.28	61.42	0	0
Delftia acidovorans CS1-4	napthalene1,2-dioxygenase	31.19	18.14	45.58	5.09	0
Polaromonas naphthalenivorans strai n CJ2.	napthalene1,2-dioxygenase	29.53	19.69	45.64	5.15	0
Ralstonia sp. Strain U2	napthalene1,2-dioxygenase	34.45	18.34	41.83	5.37	0
Terrabacter sp. Strain DBF63	phthalate dioxygenase	7.93	27.24	64.84	0	0
Alteromonas sp. SN2	ring hydroxylating dioxygenase	31.92	18.53	43.53	6.03	0
Arthrobacter phenanth renivorans Sphe 3	ring hydroxylating dioxygenase	15.35	20.12	64.52	0	0
polycyclovorans algicola TG408	ring hydroxylating dioxygenase	34.88	18.76	41.72	4.64	0

Table 4: secondary structure of all enzymes encoded by pahE gene

Bacteria	*PahE*	Helix %	Extended strands%	Coils %	Turns %	Others%
Alcaligenes	Hydratase-aldolase	46.99	10.54	35.84	0	0
Alteromonas	Dihydrodipicholinate synthetase	47.43	10.88	33.8	7.85	0
Arthrobacter	Dihydrodipicholinate synthetase	50	4.49	45.51	0	0
Burkholderia	Hydratase-aldolase	49.7	10.3	32.73	7.27	0
Citreicella	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	54.81	5.45	39.74	0	0
Delfita	Dihydrodipicholinate synthetase	47.59	10.54	35.24	6.63	0
Polaromonas	Aldolase	47.37	9.6	35.6	7.43	0
Polycyclovorans	Dihydrodipicholinate synthetase	49.55	9.97	33.53	6.95	0
Ralstonia	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	46.09	10.72	36.23	6.96	0
Terrabacter	None	none	none	none	none	none

5.2.2. Tertiary Structures:

Protein's tertiary structure is formed by secondary structure elements and maintained by interactions between their side chains. The primary structure dictates the tertiary structure and function of the molecule [53]. The properties of amino acids highly influence the shape of a protein due to their side chains [53]. A great deal of molecular interaction occurs when a protein folds into its 3D conformation [53]. Most significant of them are hydrophobic interactions between amino acids with non-polar groups which cluster at the core, stabilized by weak van der Waal forces. The shape of the protein is also affected by hydrogen bonding and ionic interactions between amino acids with polar charged groups [53].

It is the arrangement of atoms within a protein structure that determines its biological function. It could be the arrangement of catalytic residues within an active site or the way a protein interacts with other proteins to shape or regulate their structure, allowing us to formulate hypotheses about how it can be manipulated, controlled, or modified [52].

All the 3D structure of the proteins is modelled with 99 % coverage and 100% confidence. [Table 5] consists of the list of model they used and the sequence identity they found after performing the sequence identity with known protein sequences. Models of all proteins [see figure 10,11] were found to have sequence identities greater than 30% in accordance with Baker and Sali's homology modelling rule [46]. However, the protein models need to be further validated for docking purposes.

Table 5: 3D structures of all enzymes with model name. confidence and identity value

PHYRE2								
Strains	**Enzymes(pahAC)**	**Model**	**Confidence**	**Identity %**	**Enzyme (pahE)**	**Model**	**Confidence**	**Identity %**
Alcaligenes faecalis AFK2	dioxygenase large subunit	c2hmnA	100%	48	hydratase-aldolase	c6daoB	100%	57
Alteromonas sp. SN2	ring hydroxylating dioxygenase	c2hmnA	100%		dihydrodipicholinate synthetase	c6daoB	100%	60
Arthrobacter phenanthrenivorans Sphe 3	ring hydroxylating dioxygenase	c2b1xE	100%	78	dihydrodipicholinate synthetase	c6daqA	100%	40
Burkholderia sartisoli rp007	dioxygenase alpha subunit	c2hmnA	100%	57	hydratase-aldolase	c6daqA	100%	74
Citreicella aestuarii 357	IPB-dioxygenase, ISP large subunit	c1uljA	100%	58	trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	c6daoB	100%	58
Delftia acidovorans CS1-4	napthalene 1,2-dioxygenase	c2hmnA	100%	49	dihydrodipicholinate synthetase	c6daoB	100%	57
Polaromonas naphthalenivorans strain CJ2.	napthalene 1,2-dioxygenase	c2hmnA	100%	89	aldolase	c6daoB	100%	85
polycyclovorans algicola TG408	ring hydroxylating dioxygenase	c2hmnA	100%	51	dihydrodipicholinate synthetase	c6daoB	100%	58
Ralstonia sp. Strain U2	napthalene 1,2-dioxygenase	c2hmnA	100%	89	trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	c6daoB	100%	88
Terrabacter sp. Strain DBF63	phthalate dioxygenase	c2b1xE	100%	44	none			

5.2.3. Validation:

The macromolecular structure validation process involves comparing 3-dimensional atomic models of large biological molecules to statistically evaluated structural models offered by different structural databases [47]. The majority of protein structures (> 90%) submitted to PDB each year are homologous to previously established structures [48]. As a result, validation tools provide estimates about the quality of the model protein so that it can be used for further analysis.

We validate our protein models using two software packages. All proteins examined by ERRACT software had an "overall model quality" greater than 50 [44], indicating that the protein models we are interested in are of good quality**.** For our Ramachandran plots, the amino acids in allowed regions are not greater than 90%, as this is a characteristic of high-quality models [16]. But since our docking analysis is based upon hypothetically modeled proteins and in many cases of research, less than 90% of proteins are used for in-silico research purposes, we used these models.

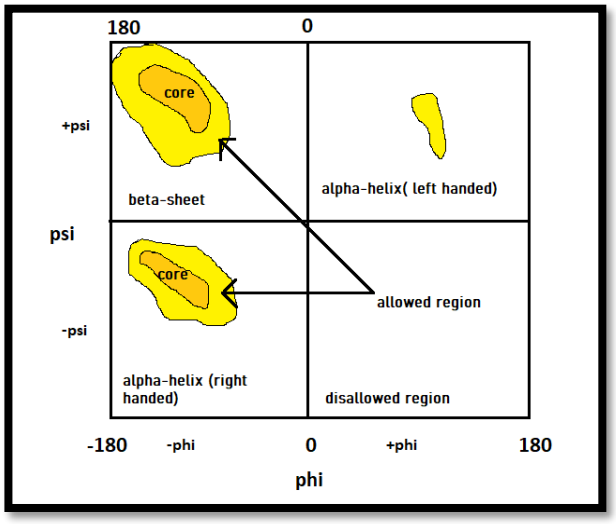

Figure 6: Ramachandran plot: The Ramachandran plot is a convenient way to visualize a protein's secondary structure. It indicates the torsional angle between amino acids phi and psi to predict a peptide's conformation The angle spectrum in each axis ranges from −180° to +180°. Atoms are treated as spheres whose dimensions are proportional to their van der Waals radii. An angle that results in the spheres colliding is sterically unfavorable; therefore, such conformations are not permitted. Marked "Core" corresponds to conformations that have no steric hindrance. Yellow areas labelled "Allowed" correspond to conformations that are possible with slightly shorter van derWaals radius values. If the atoms could come closer together, these conformations would be possible. A white area indicates
sterically un-favorable conformations [source: reference 16].

5.3. In-Silico Characterisation of Protein Molecules

Any protein's structural and molecular characterization is important in order to understand its function, properties, and parameters for its isolation and manufacturing. There are a number of factors that can be predicted by in-silico methods, including physiochemical properties, domain analysis, and the presence of its signal peptide.

5.3.1. Presence of signal peptide:

N-terminal region of proteins contains short signal peptides (SPs) sequences that convey information about protein secretion. They behave as *"postal addresses"* by targeting proteins to be secreted or transferred to other specific organelles within a cell.. Signal peptides are important for predicting where proteins are transported [24]. All the enzymesunder study have shown **no** presence of any type of signal peptide which means that theyare non-secretory and may stay in the cytosol permanently.

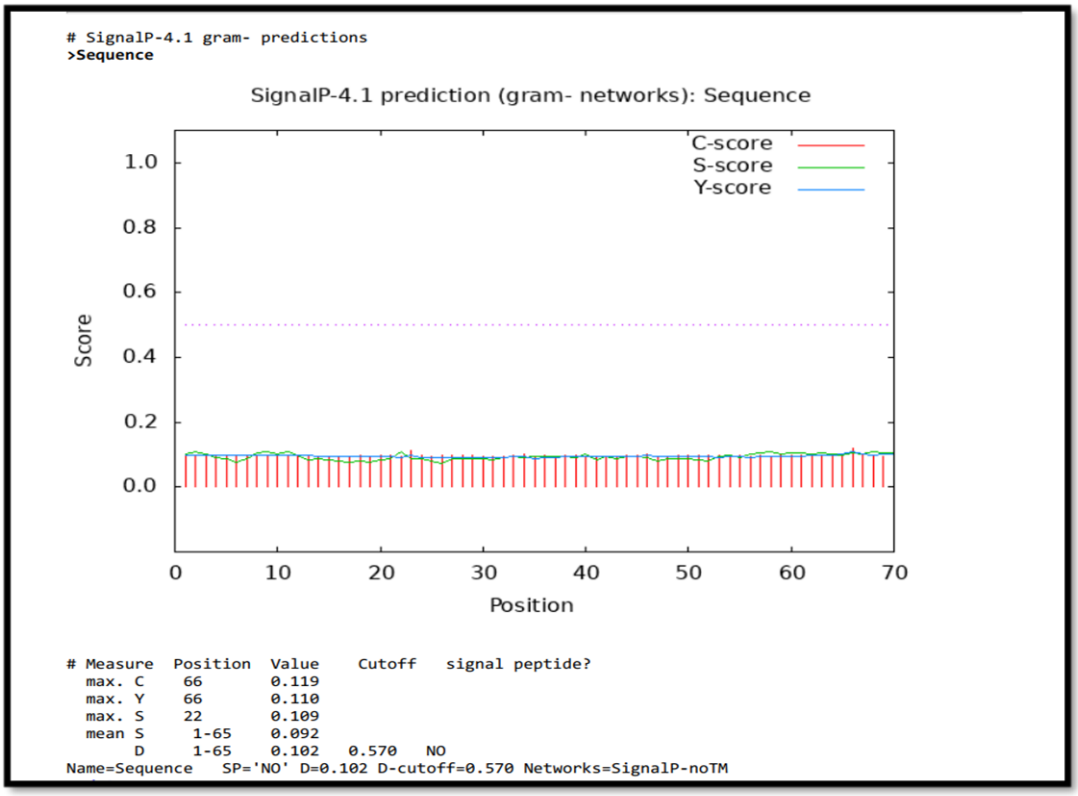

Figure 7: signalP result of Alcaligenes faecalis AFK2 for dioxygenase large subunit: The C-, S-, Y-score is below the threshold value i.e. 0.570 indicating presence of no signal peptide in the amino acid sequence of the protein.

5.3.2. Functional domains:

Three-dimensional proteins can be organized into domains that are independent folding units consisting of 100-200 residues [45,50]. Usually, long proteins contain more than one domain, each with its own specific function. When a protein possesses a sequence of domain units, its function, its localization within the cell, and its interactions are predicted. Since domains are recurring patterns in proteins, assigning a function to a domain family can reveal the function of the proteins containing this domain, which makes automatic function prediction possible(l).

Functional domains are predicted via phyre2 and Pfam servers[see table 6]. All enzymes in our study secreted by the *pahAc* gene are members of the same family of ring-hydroxylating alpha subunits. A large group of multicomponent bacterial enzymes, ring-hydroxylating dioxygenases, catalyse one of the first steps in oxidative degradation [57]. This family of enzymes have diverse substrate specificities and protein sequences. Each enzyme contains hydroxylase and electron transfer components. The hydroxylase component of the enzyme consists of two domains - the alpha subunit domain and the beta subunit domain, whose lengths vary among organisms. In addition, the alpha subunit contains a conserved C-terminal domain that functions as a catalytic domain [58]. The Rieske subunit is anchored to a 2Fe-2S cluster at its C-terminal region [53,54,55,56]. The Rieske subunit is responsible for binding both a ubiquinol anion and a plastoquinol anion, which is then used to transfer electrons to a 2Fe-2S cluster and releases them to the cytochrome c or cytochrome f heme iron. [51,52].

All of the enzymes have a similar length of their Rieseke domain and their catalytic domain except *Terrabacter sp.*, which has a larger Rieseke domain and a smaller catalytic domain. All enzymes that are secreted from the *pahE* gene in all organisms belong to the dihydrodipicolinate synthetase family with a similar domain length. Dihydrodipicolinate synthase (DHDPS) plays an active role in lysine biosynthesis. Using a Schiff base reaction between pyruvate and lysine residues, it catalyses the aldol condensation of L-aspartate-beta- semi aldehyde and pyruvate to dihydropicolinic acid. It consists of a homo tetramer of dimers. The dihydrodipicolinate synthase enzyme belongs to a family of pyruvate-dependent aldolase which catalyze different reactions in different metabolic pathways by using the same catalytic step. [59,60,61,62].

Table 6: functional domains of all proteins by PFAM

Strain	*PahAc*	*PahE*	Pfam analysis	
Alcaligenes faecalis AFK2	Dioxygenase large subunit	Hydratase-aldolase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Alteromonas sp. SN2	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Arthrobacter phenanthrenivorans Sphe 3	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Burkholderia sartisoli rp007	Dioxygenase alpha subunit	Hydratase-aldolase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Citreicella aestuarii 357	IPB-dioxygenase, ISP large subunit	Trans-o-hydroxybenzyliden epyruvate hydratase-aldolase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Delftia acidovorans CS1-4	Napthalene1, 2-dioxygenase	Dihydrodipicholinate synthetase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Polaromonas naphthalenivorans strain CJ2.	Napthalene1, 2-dioxygenase	Aldolase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Polycyclovorans algicola TG408	Ring hydroxylating dioxygenase	Dihydrodipicholinate synthetase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Ralstonia sp. Strain U2	Napthalene1, 2-dioxygenase	Trans-o-hydroxybenzyliden epyruvate hydratase-aldolase	Rhd alpha subunit (catalytic domain)	Dihydrodipicholinate synthetase
Terrabacter sp. Strain DBF63	Pthalate dioxygenase	None	Big rieske Rieske and small rhd alpha catalytic domain	

5.3.3. Physio-chemical properties:

According to Gasteiger et al., the physiochemical properties of all enzymes were analysed using PROTPARAM. The physicochemical properties are listed in [Table7]. Analyzing the isoelectric point (theoretical pI) and molecular weight of amino acid sequences is vital as these variables help predict the approximate area of a2D gel that can be used to detect the protein of interest. An instability index indicates how stable a protein is in a test tube. A Protein's Instability index over 40 is likely to be unstable in the test tube. The enzymes tested, however, have proved to be stable. An aliphatic index (AI) is calculated by taking into account the relative volume occupied by amino acid side chains. A higher AI value indicates that the protein is more stable. According to our analysis, all proteins encoded by the *PahE* gene are in the same range of stability except for dihydrodipicolinate synthetase from *Arthrobacter sphe3* (87.96). This indicates it is the most stable enzyme for a wide range of temperatures, while all proteins encoded by the *PahAc* gene are also in the same range. The GRAVY values of all the proteins are negative in the range of -0.1 to -0.4 indicating that they are water- soluble. It is estimated that all proteins have a half-life of 30 hours which is longer than 10 hours, which reinforces the stability of proteins during purification.

Table 7: secondary structures of all enzymes under study

Strain	Enzymes	Aa	Mw	Pi	Total (-) residues	Total (+) residues	Ext. co	II	AI	GRAVY
Alcaligenes faecalis afk2	Dioxygenase large subunit	437	48278.61	5.61	52	44	78630/ 78380	28.02 (stable)	77.23	-0.287
	Hydratase-aldolase	332	36906.02	5.75	40	35	71640/ 71390	32.22 (stable)	79.13	0.195
Alteromonas sp. Sn2	Ring hydroxylating dioxygenase	448	50185.96	5.28	57	39	94810/ 94310	28.07 (stable)	68.57	-0.489
	Dihydrodipicholinate synthetase	331	36819.15	5.59	38	34	60890/ 60390	27.79 (stable)	77.28	-0.184

Strain	Enzymes	Aa	Mw	Pi	Total (-) residues	Total (+) residues	Ext. co	II	AI	GRAVY
Arthrobacter phenanthrenivorans sphe 3	Ring hydroxylating dioxygenase	482	54455.44	5.44	69	48	87570/ 87320	28.71 (stable)	69.61	-0.507
	Dihydrodipicholinate synthetase	334	36899.25	4.82	46	34	55015/ 54890	38.48 (stable)	87.96	-0.1
Burkholderia sartisoli rp007	Dioxygenase alpha subunit	450	50535.84	5.91	56	46	94810/ 94310	32.50 (stable)	68.73	-0.444
	Hydratase-aldolase	330	36473.46	5.24	44	37	63160/ 62910	37.46 (stable)	78.21	-0.201
Citreicella aestuarii 357	IPB-dioxygenase, isp large subunit	451	50637.75	5.42	58	43	94685/ 94310	29.64 (stable)	68.82	-0.465
	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	312	34495.2	5.05	43	35	60640/ 60390	29.39 (stable)	79.52	-0.184
Delftia acidovorans cs1-4	Napthalene 1,2-dioxygenase	452	50214.84	5.66	55	46	90090/ 89840	23.65 (stable)	76.84	-0.323
	Dihydrodipicholinate synthetase	332	36891.99	5.75	40	35	71640/ 71390	31.89 (stable)	78.83	-0.202
Polaromonas naphthalenivorans strain cj2.	Napthalene 1,2-dioxygenase	447	49548.74	5.78	55	44	87695/ 87321	34.73 (stable)	73.56	-0.36
	Aldolase	323	35785.9	6.92	42	42	59150/ 58900	27.22 (stable)	72.91	-0.305
Polycyclovorans algicola tg408	Ring hydroxylating dioxygenase	453	50131.64	6.21	54	48	87360/ 86860	29.42 (stable)	70.91	-0.307
	Dihydrodipicholinate synthetase	331	36415.77	6.84	37	37	61795/ 61420	33.35 (stable)	79.37	-0.111

Strain	Enzymes	Aa	Mw	Pi	Total (-) residues	Total (+) residues	Ext. co	II	AI	GRAVY
Ralstonia sp . Strain u2	Napthalene 1,2-dioxygenase	447	49570.7	5.78	55	44	87695/ 87320	34.16 (stable)	72.66	-0.371
	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	345	38180.7	8.51	43	46	64650/ 64400	25.30 (stable)	75.91	-0.272
Terrabacter sp. Strain dbf63	Pthalate dioxygenase	492	55144.81	5.35	68	51	83560/ 83310	34.59 (stable)	73.39	-0.441

5.4 Docking Analysis And Visualizations:

Molecular docking is a method used to simulate interactions between ligands and proteins at an atomic level using computational methods [63]. An accurate Docking process usually involves 2 fundamental steps-a) finding out the correct ligand conformation with the protein molecule b) assessing the binding affinity of the docked protein-ligand complexes [69,70].

The binding affinity can be characterized as the strength of the interaction between a protein molecule and its ligand, normally measured by equilibrium dissociation constants (KDs), which are used to determine the order strength of bimolecular interactions. A lower KD value indicates a more specific binding affinity for the target. A larger KD value indicates weaker binding between the target molecule and ligand [71]. Each docking process generates a number of binding conformations and corresponding binding affinities. Protein-ligand complexes with lower binding affinities have better predicted conformations [31] . A summary of their docking analysis is given in table 7 and table 8. It is important to remember that prior to any docking process, the respective ligands to be used for docking need to be energy minimized.
The theory of energy minimization before docking is based on removing atom clashes in the ligand and developing a reasonable starting position before docking. Each docking provides a list of different conformations of the protein-ligand complexes and corresponding binding affinity tabulated in the result section. The best pose with the lowest binding affinity is taken for all enzymes.

5.4.1. Analysis for enzymes encoded by *pahAc* gene (table 8):

- *Arthrobacter sp.*, which is known for producing ring-hydroxylating dioxygenase, binds phenanthrene with the lowest efficiency (- 7.4 Kcal/mol) than any other enzyme in any other organism. It is *Delftia sp.* that degrades phenanthrene with the highest efficiency (-9.1 Kcal/mol).
- Despite using different enzymes, *Alteromonas sp., Alcaligenes sp., and Burkholderia sp*. degrade phenanthrene to the similar degree.
- *Alteromonas sp.* encodes the RHDalpha enzyme that degrade naphthalene, phenanthrene, anthracene, and pyrene. The docking results showed it to have the highest binding affinity for pyrene with a value of -9.5 Kcal/mol, followed by anthracene, phenanthrene, andnaphthalene.
- *Burkholderia sp.* enzyme breaks down naphthalene, phenanthrene, and anthracene. Compared to other compounds, it destroys anthracene most efficiently.
- The efficiency of degrading naphthalene by organisms that degrade it (*Alteromonas sp., Burkholderia sp., Polaromonas sp., Polycyclovorans sp., Ralstonia sp.*) do not differ significantly regardless of the enzyme produced by them except *citreicella sp.*, that use IPB dioxygenase have generated the least amount of degradation.

Table 2: docking result of all enzymes encoded by pahAc gene of all organisms using pyrx software

Strains	Enzymes *(pahAc)*	Ligand	Binding affinity (kcal/mol)
Burkholderia sartisoli RP007	Dioxygenase alpha subunit	Naphthalene, phenanthrene and anthracene	-6.8,-9,-9.1
Terrabacter sp. Strain DBF63	Pthalate dioxygenase	Fluorene	-7.5
Citreicella aestuarii 357	IPB-dioxygenase, isp large subunit	Napthalene	-5.8
Polaromonas naphthalenivorans strain CJ2.	Napthalene1,2-dioxygenase	Napthalene	-6.8
Polycyclovorans algicola TG408	Ring hydroxylating dioxygenase	Napthalene	-6.8
Ralstonia sp. Strain U2	Napthalene1,2-dioxygenase	Napthalene	-6.8
Alteromonas sp. SN2	Ring hydroxylating dioxygenase	Napthalene, phenanthrene,anthracene, pyrene	-6.5, -8.9,-8.9,-9.5
Alcaligenes faecalis AFK2	Dioxygenase large subunit	Phenanthrene	-8.3
Arthrobacter phenanthrenivorans sphe 3	Ring hydroxylating dioxygenase	Phenanthrene	-7.4
Delftia acidovorans CS1-4	Napthalene1,2-dioxygenase	Phenanthrene	-9.1

- Various experiments have shown that ring-hydroxylating dioxygenase can degrade both phenanthrene and naphthalene in different organisms. The RHD alpha enzyme in *Arthrobacter sp*. degrades phenanthrene, while in *Polycyclovorans sp.* it degrades naphthalene. In *Alteromonas sp*., the RHD alpha enzyme degrade naphthalene, phenanthrene, anthracene and pyrene. As a consequence, we can hypothesize. We tested our hypothesis by docking RHD alpha enzymes from *Arthrobacter sp. and Polycyclovorans sp.* with phenanthrene, naphthalene and pyrene molecules comparing it with *Alteromonas sp.* It was found that *Arthrobacter sp.,* which degrades phenanthrene, had the highest affinity to pyrene. Nevertheless, *Polycylovorans sp.* that degrade naphthalene can also metabolize phenanthrene, anthracene and pyrene with higher efficiency than naphthalene. It has the highest efficiency for pyrene in terms of binding affinity. In addition to what is determined by a conventional method of experimentation, an in-silico method has shown that *Arthobacter sp. and Polycylovorans sp*. can also degrade other PAH compounds using their same enzyme in respective organisms [Table 9.2] [figure 8.3-8.8].

- The napthalene1,2 dioxygenase enzyme was also studied in a similar way. Naphthalene dioxygenase is capable of degrading naphthalene in *Ralstonia sp., Polaromonas sp.,* and phenanthrene in *Delftia sp*. Thus, we can assume that naphthalene dioxygenase is capable of degrading both naphthalene and phenanthrene in these three organisms. We observed that [table 9.1][figure 8.1-8.2] -
  - While the *Deftia sp.* has a naphthalene dioxygenase enzyme, phenanthrene is what it degrades much more effectively compared to naphthalene
  - Compared to the other 2 organisms, *Delftia sp.* is more efficient at degrading naphthalene and phenanthrene.
  - In addition to degrading naphthalene, *Polaromonas sp. and Ralstonia sp.* also degrade phenanthrene more effectively.

Table 9.1 : DOCKING RESULT USING PYRX: WHEN NAPTHALENE 1,2 DIOXYGENASE FROM 3 ORGANIMS DOCKED WITH A: NAPHTHALENE B: PHENANTHRENE. The black highlighted letters mean that are already found to have shown degradation, the red letters represent that we have hypothesized to show degradation

A: NAPHTHALENE

Enzymes	Ligand	Species	Binding affinity
Napthalene 1,2 dioxygenase	Napthalene	***Ralstonia sp. U2***	-6.8
Napthalene 1,2 dioxygenase	Napthalene	***Polaromonas naphthalenivorans CJ2***	-6.8
Napthalene 1,2 dioxygenase	Napthalene	***Delftia acidovorans CS1-4***	-7.1

B: PHENANTHRENE

Enzymes	Ligand	Species	Binding affinity
Napthalene 1,2 dioxygenase	Phenanthrene	***Delftia acidovorans CS1-4***	-9.1
Napthalene 1,2 dioxygenase	Phenanthrene	***Ralstonia sp. U2***	-8.7
Napthalene 1,2 dioxygenase	Phenanthrene	***Polaromonas naphthalenivorans CJ2***	-8

Table 9.2: docking result using PYRX: A: when RHD alpha enzymes from 3 organisms docked with A: naphthalene B: Phenanthrene C: pyrene d: Anthracene. The black highlighted letters mean that are already found to have shown degradation, the red letters represent that we have hypothesized to show degradation.

A: NAPHTHALENE

Enzymes	Ligand	Species	Binding affinity
RHD alpha	Napthalene	***Arthrobacter phenanthrenivorans sphe 3***	-8.3
RHD alpha	Napthalene	***Alteromonas sp. SN2***	-6.5
RHDalpha	Napthalene	***Polycyclovorans algicola TG408***	-6.8

B: PHENANTHRENE

Enzymes	Ligand	Species	Binding affinity
RHDalpha	Phenanthrene	**Arthrobacter phenanthrenivorans sphe 3**	-7.4
RHDalpha	Phenanthrene	**Alteromonas sp. SN2**	-8.9
RHDalpha	Phenanthrene	**Polycyclovorans algicola TG408**	-8.9

C: PYRENE

Enzymes	Ligand	Species	Binding affinity
RHDalpha	Pyrene	**Alteromonas sp. SN2**	-9.5
RHDalpha	Pyrene	**Arthrobacter phenanthrenivorans sphe 3**	-11.8
RHDalpha	Pyrene	**Polycyclovorans algicola TG408**	-9.5

D: ANTHRACENE

Enzymes	Ligand	Species	Binding affinity
RHDalpha	**Anthracene**	Alteromonas sp.SN2	**-8.9**
RHDalpha	Anthracene	**Polycyclovorans algicola TG408**	-8.7
RHDalpha	Anthracene	**Arthrobacter phenanthrenivorans sphe2**	-7.3

5.4.2. Analysis for enzymes encoded by pahE gene (table 10):

All enzymes encoded by the *pahE* gene bind with the ligand pyruvate [figure 9.1-9.2] to carry on the metabolic degradation of poly-aromatic compounds in respective organisms. Our docking analysis shows how that: -

- Secreting enzymes have similar binding efficiencies to pyruvate, except in *Alteromonas sp.*, with a binding affinity of -3.8 Kcal/mol, it is the least efficient ligand binder.
- Table 10 shows that hydratase-aldolase enzymes from *Polaromonas sp., and Citreicella sp.*, have the lowest binding affinity. A low binding affinity means more stability and better ligand binding.

Visualizing all docked protein-ligand molecules with Discovery Studio has shown that no pocket molecule matches those that use the same protein enzyme for the same ligand molecule. As an example, it has been found that when degrading phenanthrene with the same enzymes such as RHD or dioxygenase, none of them showed identical binding to the pocket or identical interacting atoms. Similarly, hydratase-aldolase enzymes carried by all organisms that encode it display no similar interacting atoms or pocket atoms among themselves, either. In some cases, the molecular types of the interacting atom may have been similar but not identical even though they use the same enzymes, and yield the same binding affinity for the same ligand. It is therefore impossible to predict the impact of interacting atoms.

Table 10: docking result of all enzymes encoded by pahE gene from all organisms using PYRX software

Bacteria	Enzymes (pahE)	Ligand	Binding affinity (Kcal/mol)
Polaromonas	Aldolase	Pyruvate	-4.3
Alteromonas	Dihydrodipicholinate synthetase	Pyruvate	-3.8
Arthrobacter	Dihydrodipicholinate synthetase	Pyruvate	-4.1
Delfita	Dihydrodipicholinate synthetase	Pyruvate	-4.2
Polycyclovorans	Dihydrodipicholinate synthetase	Pyruvate	-4.1
Alcaligenes	Hydratase-aldolase	Pyruvate	-4
Burkholderia	Hydratase-aldolase	Pyruvate	-4.1
Citreicella	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	Pyruvate	-4.3
Ralstonia	Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase	Pyruvate	-4.2

CHAPTER 6: CONCLUSION

We can acquire much more reliable information for various complicated problems when we combine approaches from different domains. Two of these approaches are bioinformatics and bioremediation, which are used in this study. We have successfully characterised degrading enzyme molecules from 10 new organisms found capable of PAH bioremediation. Their degradation capability is also analysed via docking studies. We have seen, Arthrobacter *phenanthrenivorans Sphe 3* degrades not only phenanthrene but also napthalene and pyrene with highest efficiency. *Delftia acidovorans CS1-4* has the highest efficiency of degradation for phenanthrene, and *Alteromonas sp. SN2* has the highest efficiency for anthracene. All organisms have the same ability to degrade naphthalene. Even though these degradation abilities haven't been established by experimental means, in-silico techniques have shown that they degrade. A similar pattern of binding molecules was not observed in any protein-ligand complex

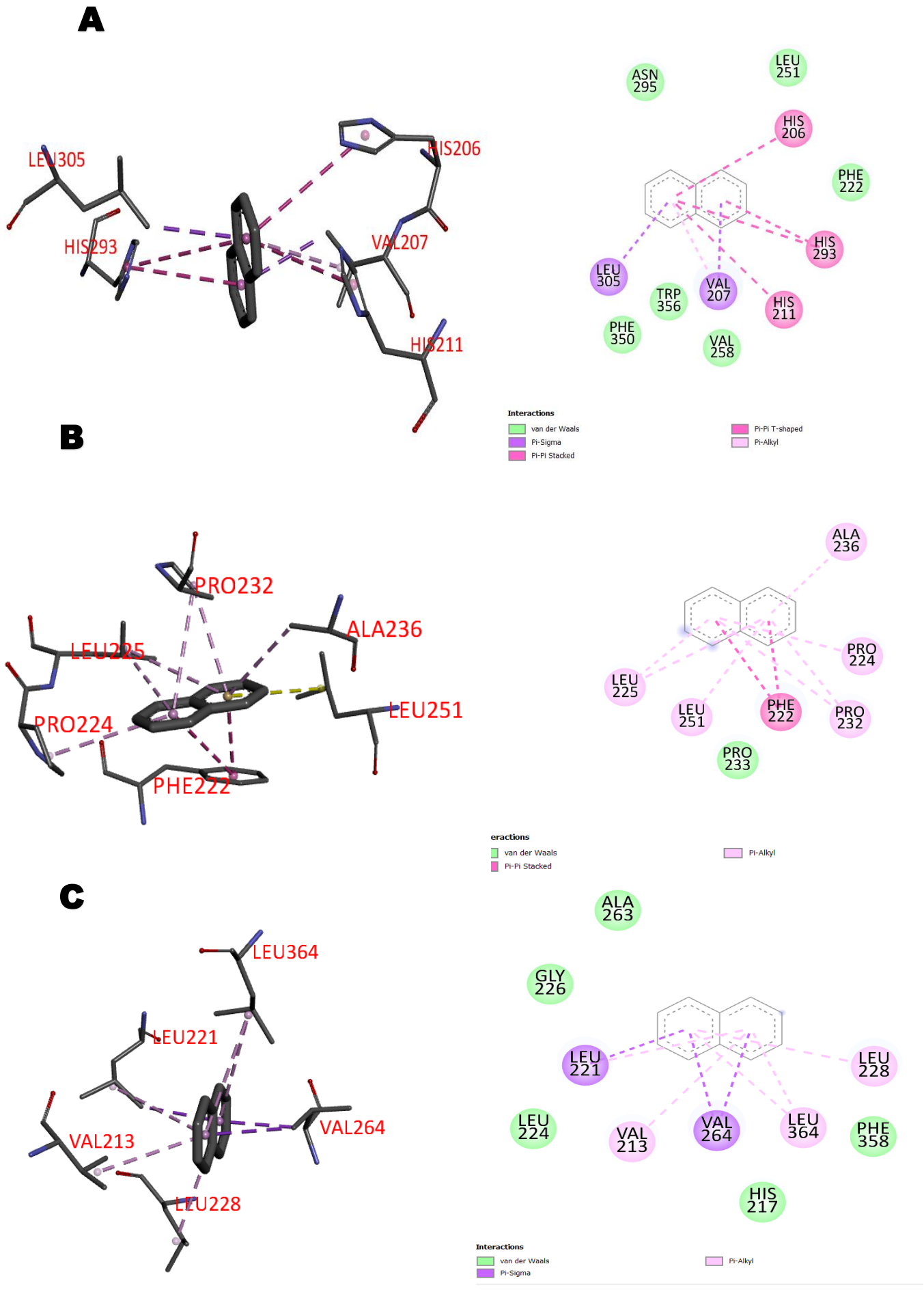

Figure 8.1: 3D and 2D diagram of naphthalene 1,2 dioxygenase enzymes bounded to naphthalene ligand molecule: A: *Ralstonia* sp. B: *Polaromonas sp.* C: *Delftia sp.* The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

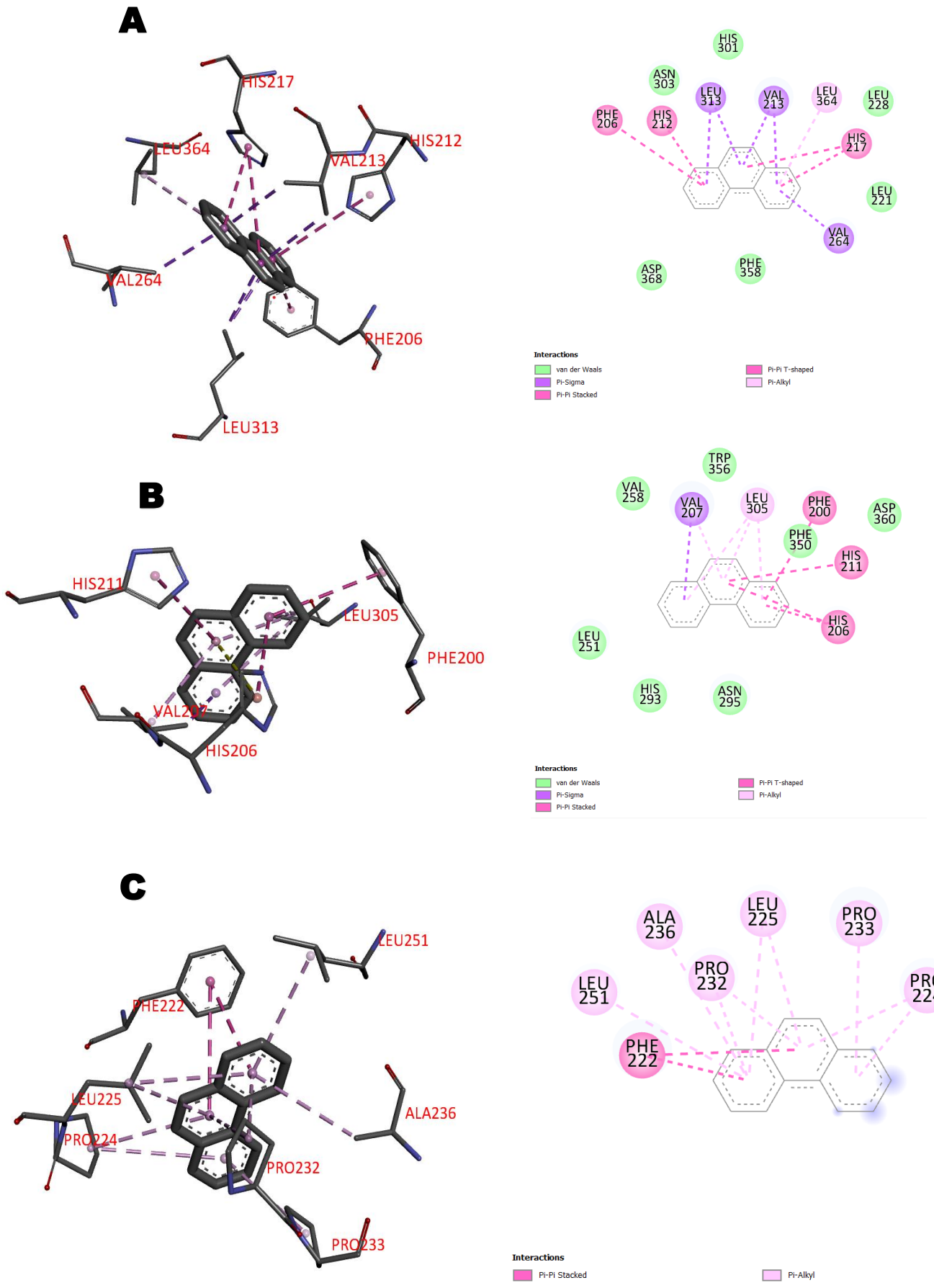

Figure 8.2: 3D and 2D diagram of naphthalene 1,2 dioxygenase enzymes bounded to phenanthrene ligand molecule: A: *Delftia* sp. B: *Ralstonia sp.* C: *Polaromonas sp.* The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

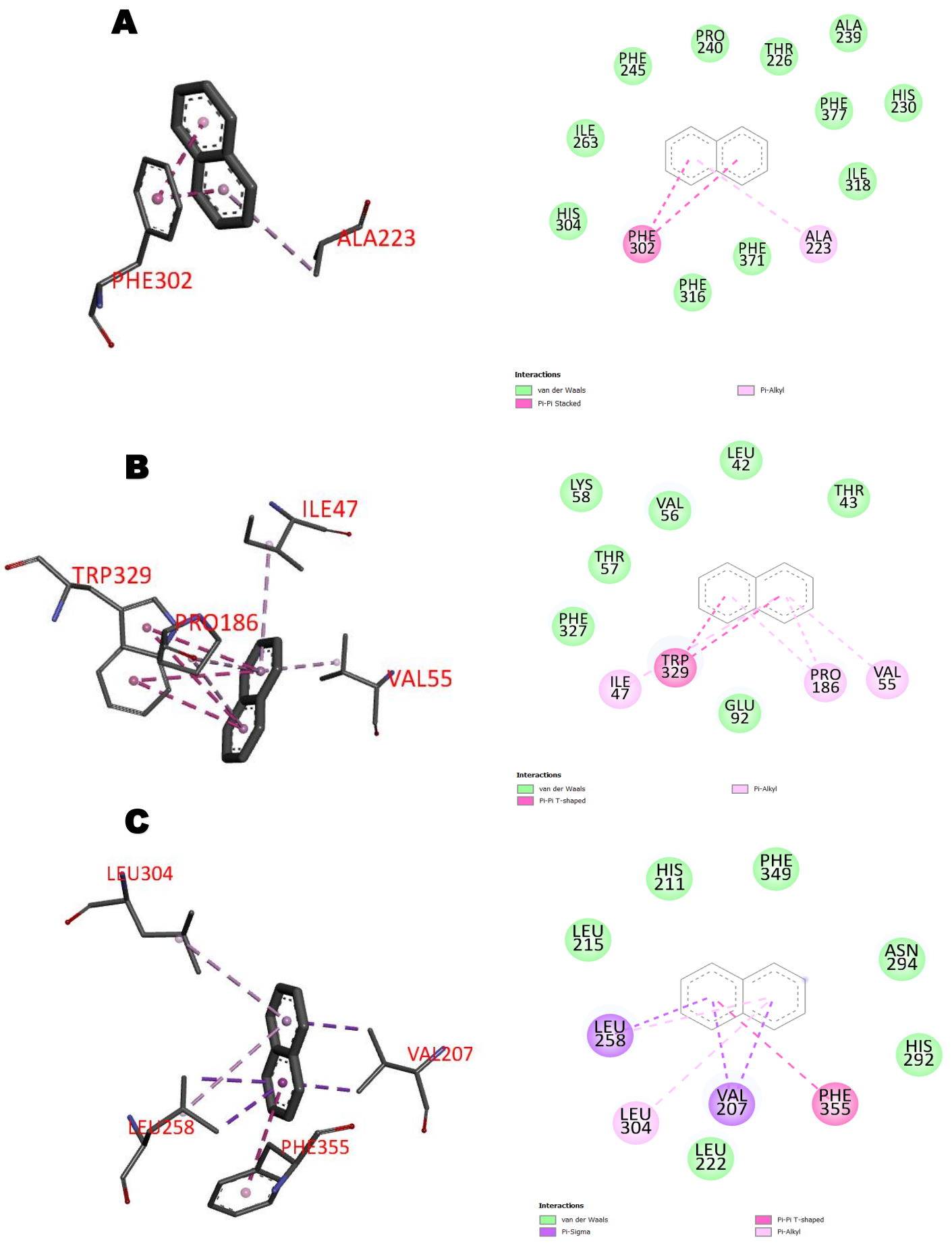

Figure 8.3 : 3D and 2D diagram of ring-hydroxylating dioxygenase enzymes bounded to naphthalene ligand molecule: A: Arthrobacter sp. B: Alteromonas sp. C: Polaromonas sp. The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

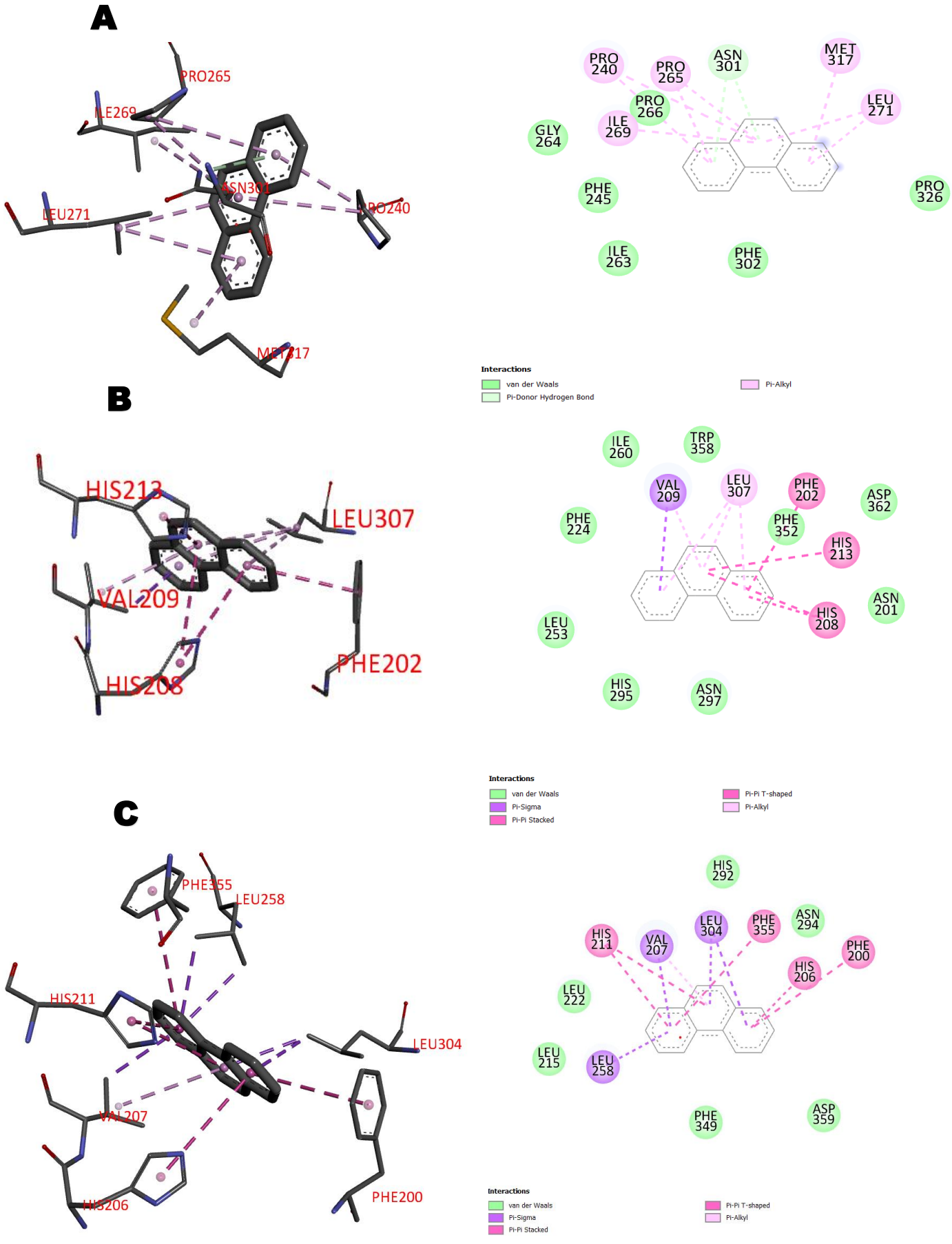

Figure 8.4 : 3D and 2D diagram of ring-hydroxylating dioxygenase enzymes bounded to Phenanthrene ligand molecule: A: Arthrobacter sp. B: Alteromonas sp. C: Polaromonas sp. The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

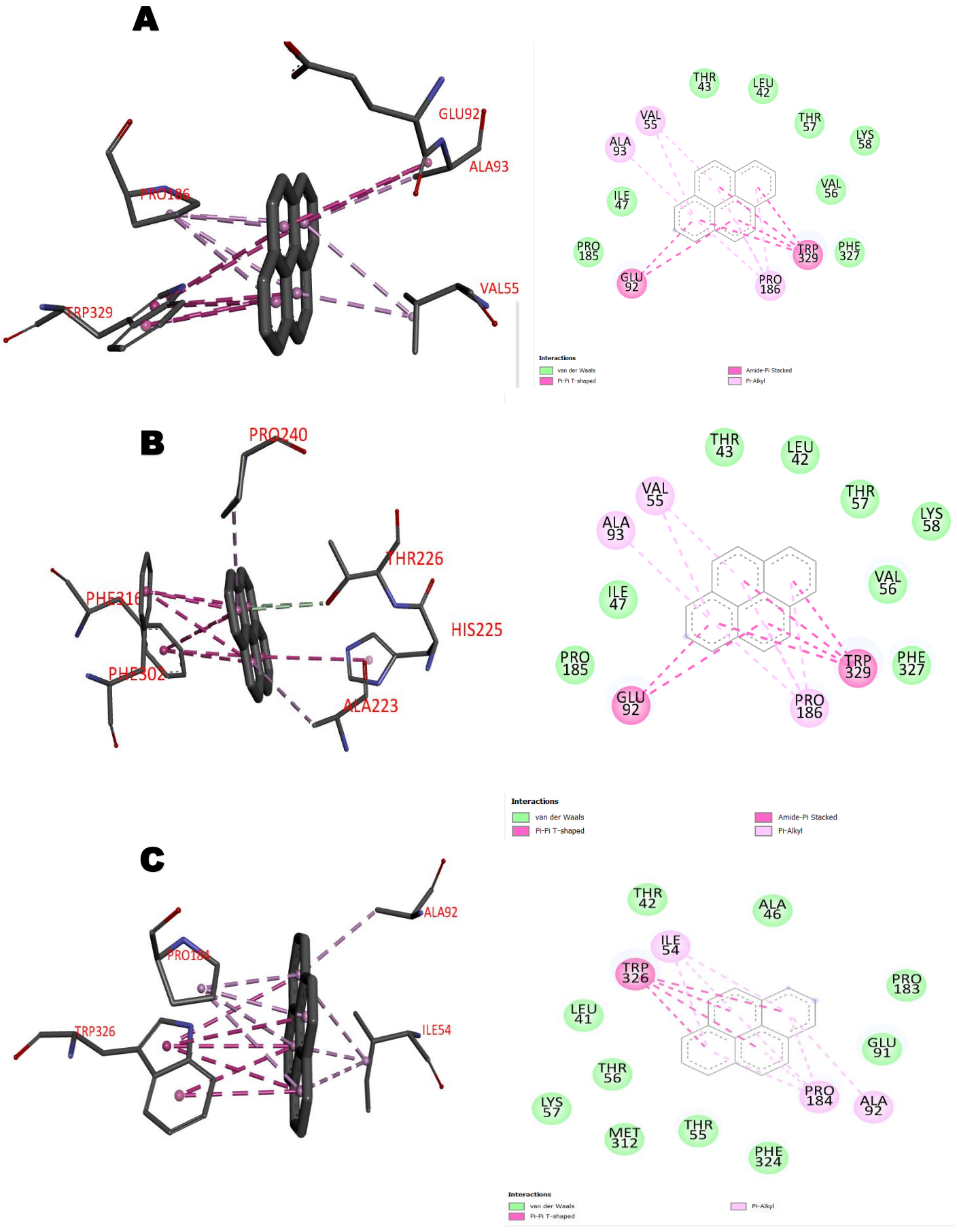

Figure 8.5 : 3D and 2D diagram of ring-hydroxylating dioxygenase enzymes bounded to Pyrene ligand molecule: A: Alteromonas sp. B: Arthrobacter sp. C: Polycyclovorans sp. The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

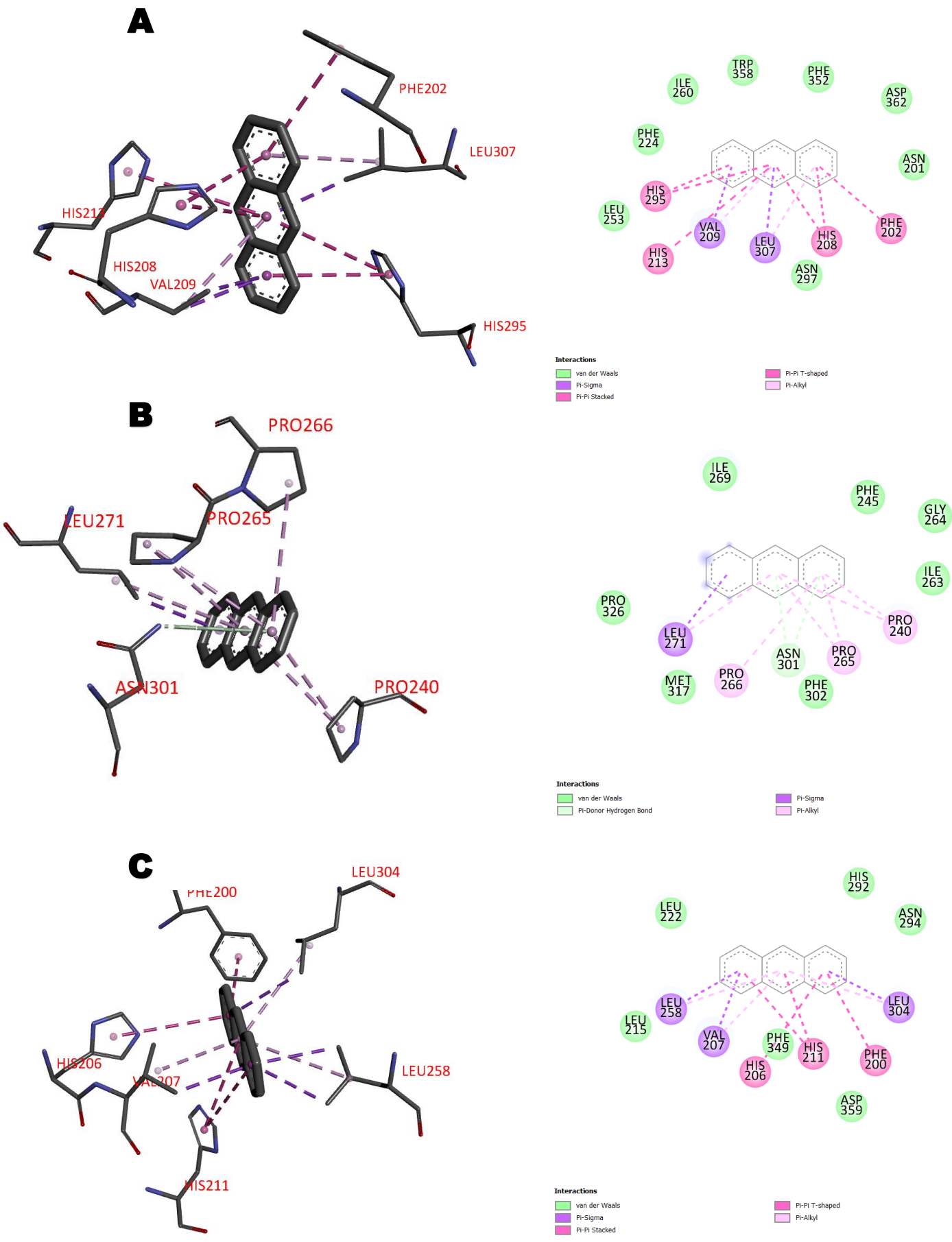

Figure 8.6 : 3D and 2D diagram of ring-hydroxylating dioxygenase enzymes bounded to Anthracene ligand molecule: A: Alteromonas sp. B: Arthrobacter sp. C: Polycyclovorans sp. The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

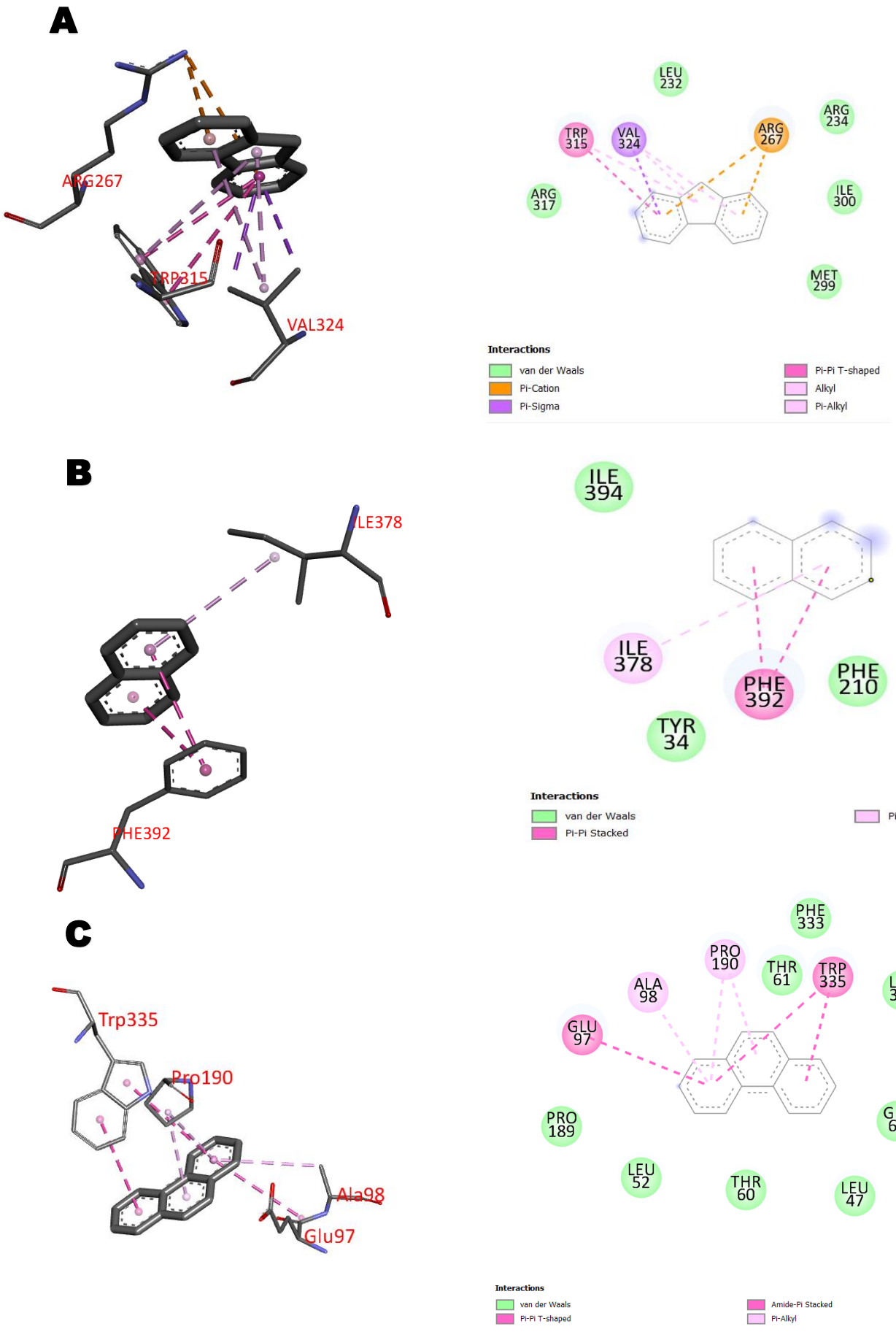

Figure 8.7 : 3D and 2D diagram of molecule: A: Terrabacter sp. B: Citreicella sp. C: Alcaligenes sp. bound to fluorene, naphthalene, and phenanthrene The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

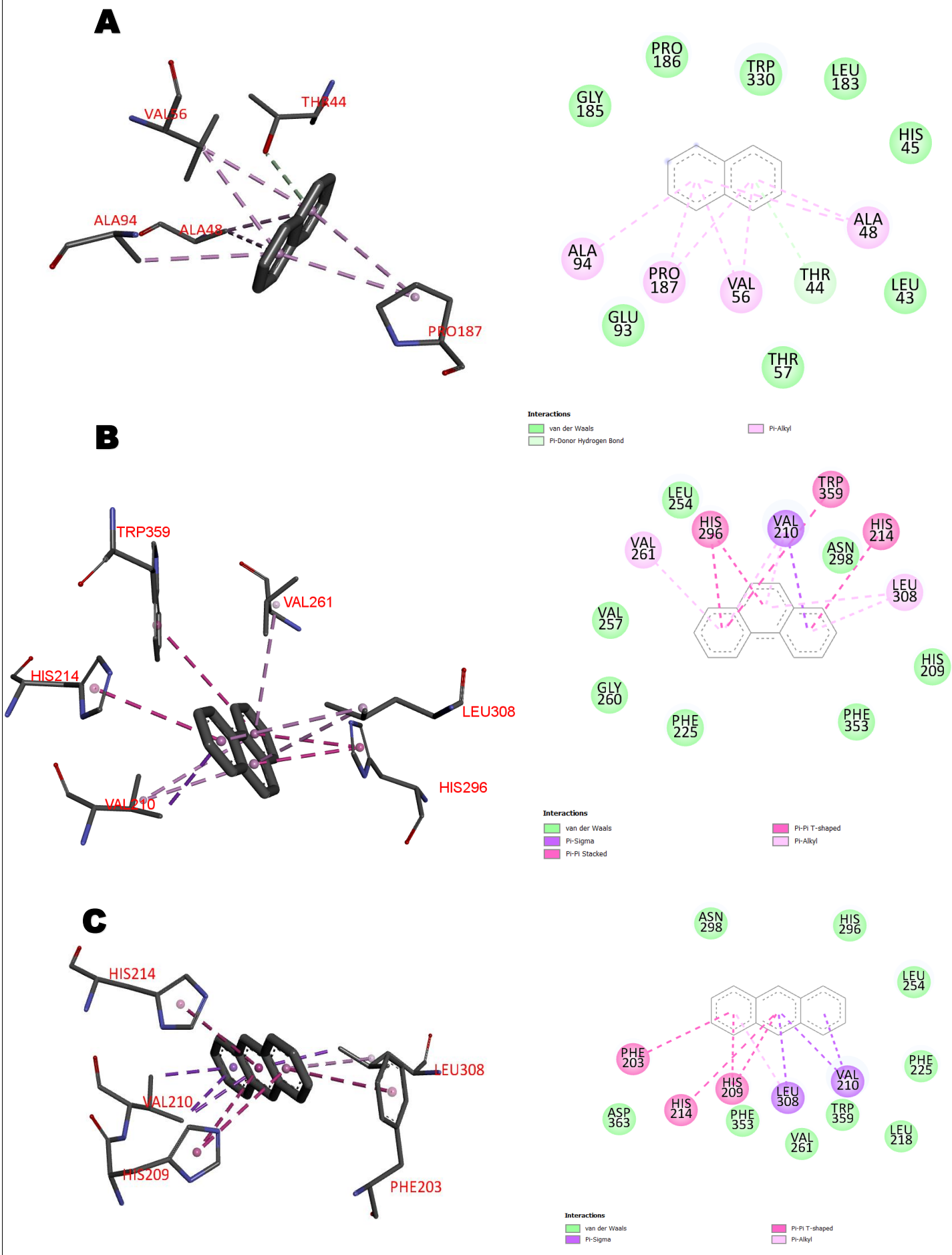

Figure 8.8 : 3D and 2D diagram of molecule: A:Naphthalene. B: phenanthrene. C: anthracene bound to dioxygenase enzyme of *Burkholderia sp*.The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

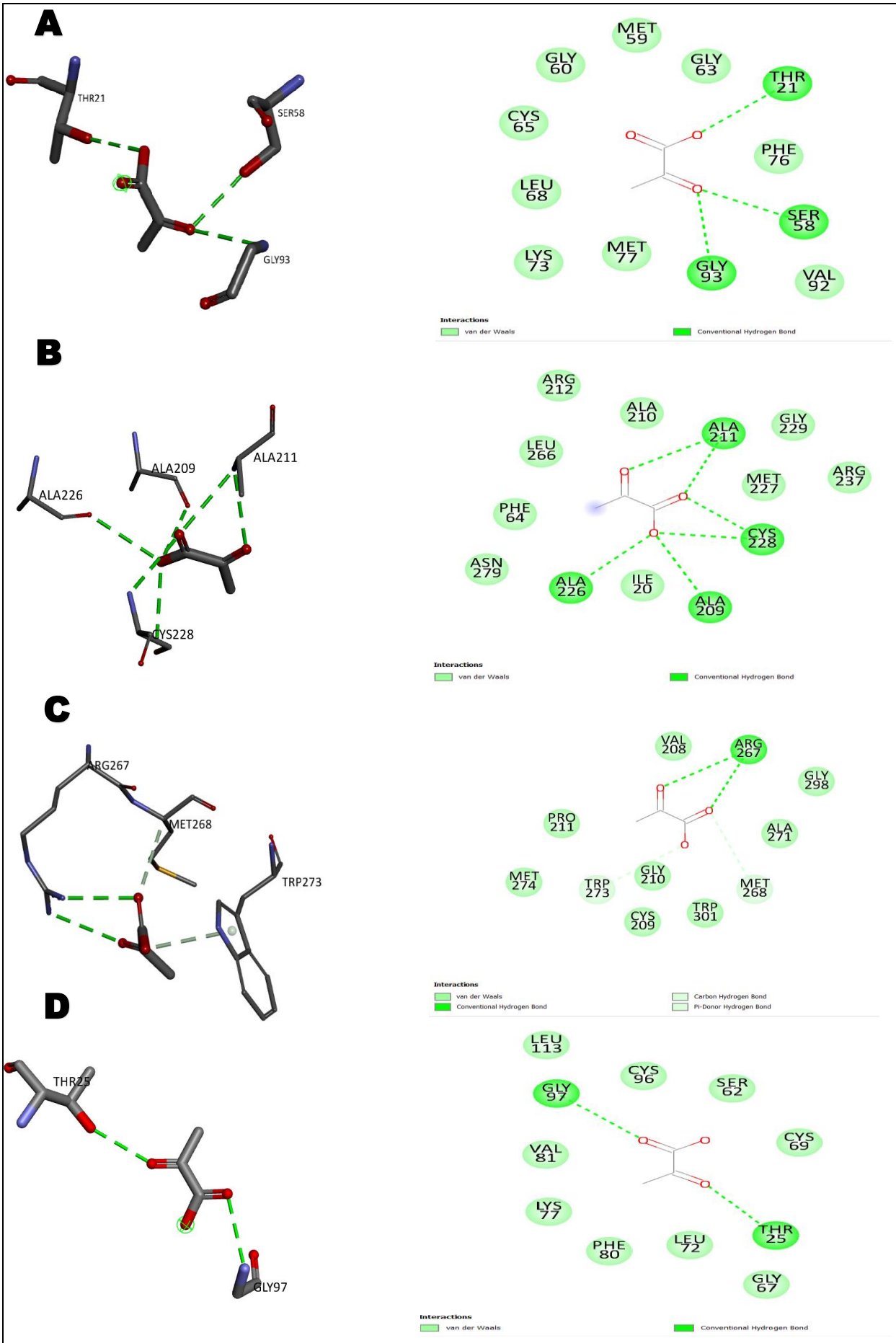

Figure 9.1: 3D and 2D diagram of hydratase-aldolase enzymes bounded to pyruvate ligand molecule: A: *Alcaligenes* sp. B: *Burkholderia sp.* C: *Citreicella sp.* D: *Ralstonia sp*. The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

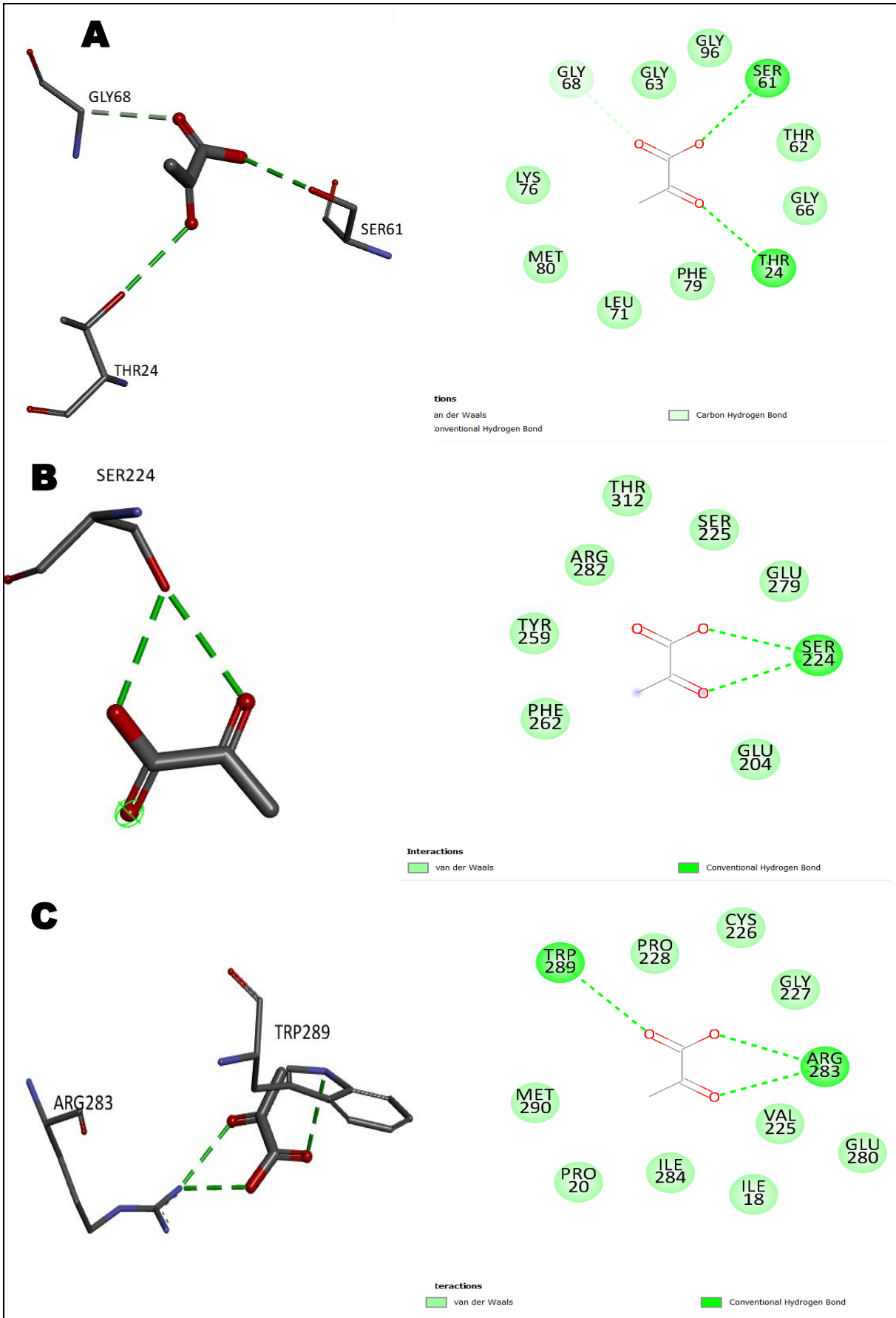

Figure 9.2: 3D and 2D diagram of dihydrodipicolinate synthetase enzymes bounded to pyruvate ligand molecule: A: *Alteromonas* sp. B: *Arthrobacter sp.* C: *Delftia sp.* The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

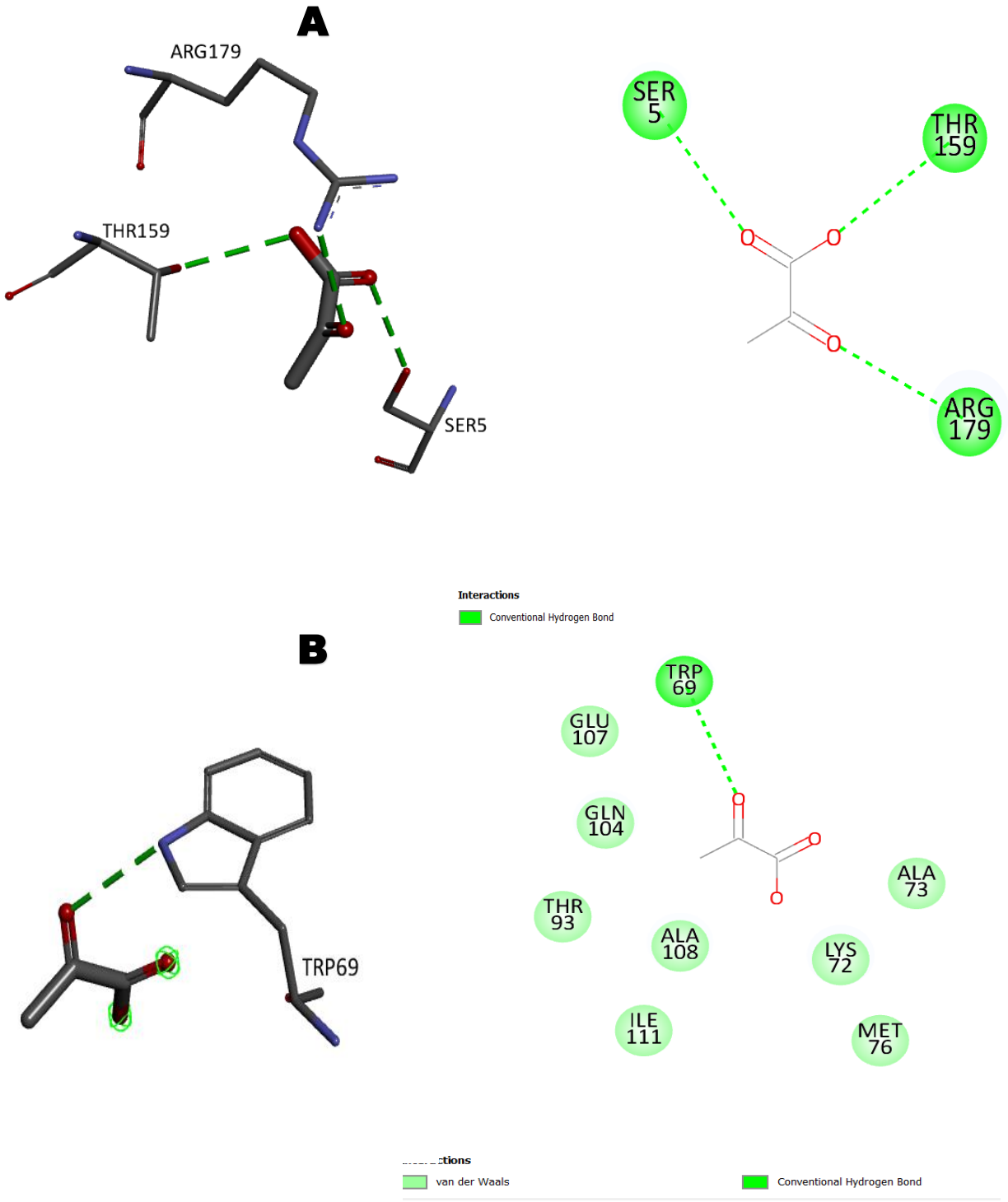

Figure 9.3: 3D and 2D diagram of dihydrodipicolinate synthetase enzymes bounded to pyruvate ligand molecule: A: *Polaromonas* sp. B: *Polycylovorans sp.* The interacting amino acid and the types of bond formed mentioned below each 2D diagrams.

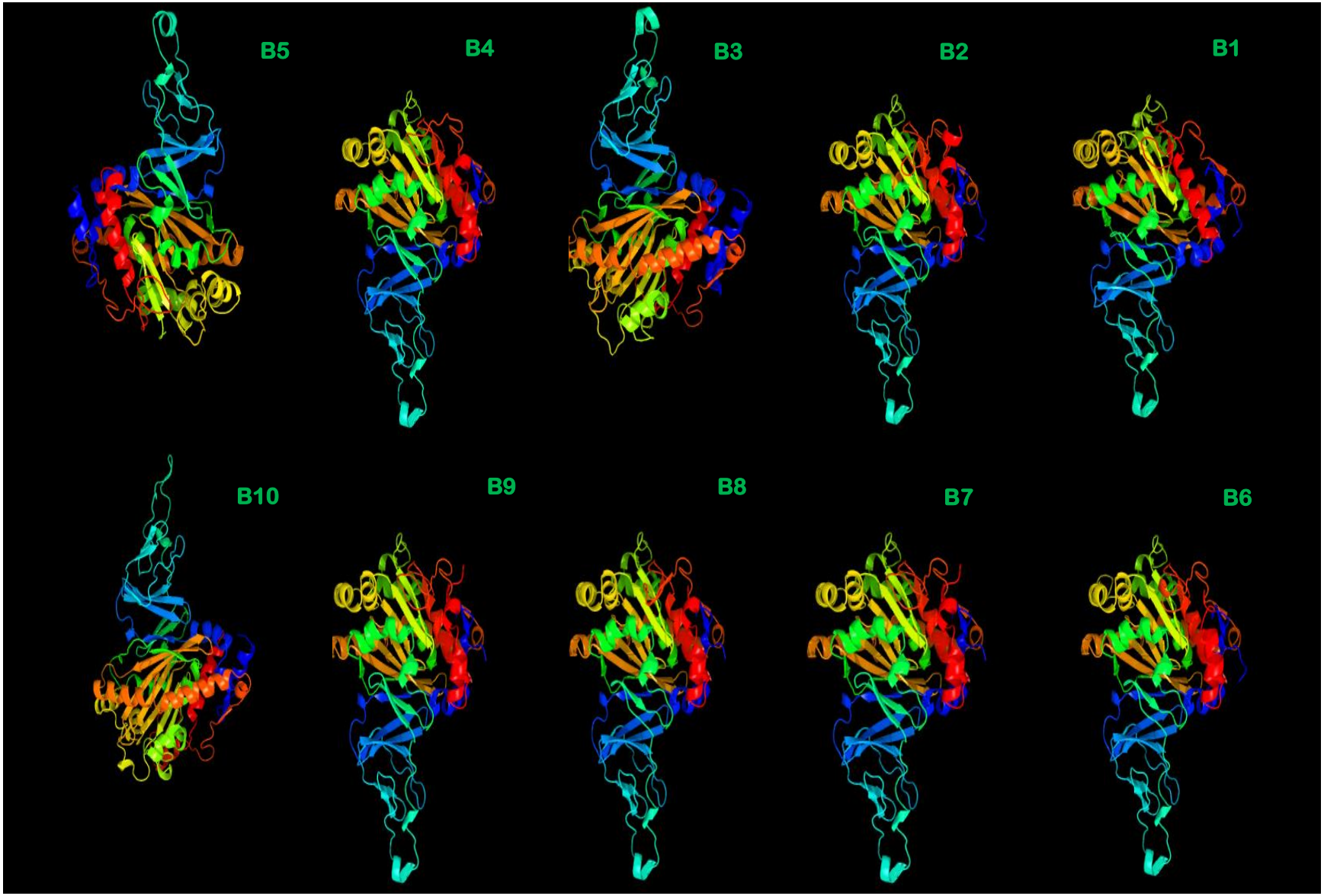

Figure 10 : A: 3D structures of pahAc Enzymes: 1: Alcaligenes 2. Alteromonas 3. Arthrobacter 4. Burkholderia 5. Citreicella 6: Delftia 7: Polaromonas 8: polycyclovorans 9. Ralstonia 10. Terrabacter

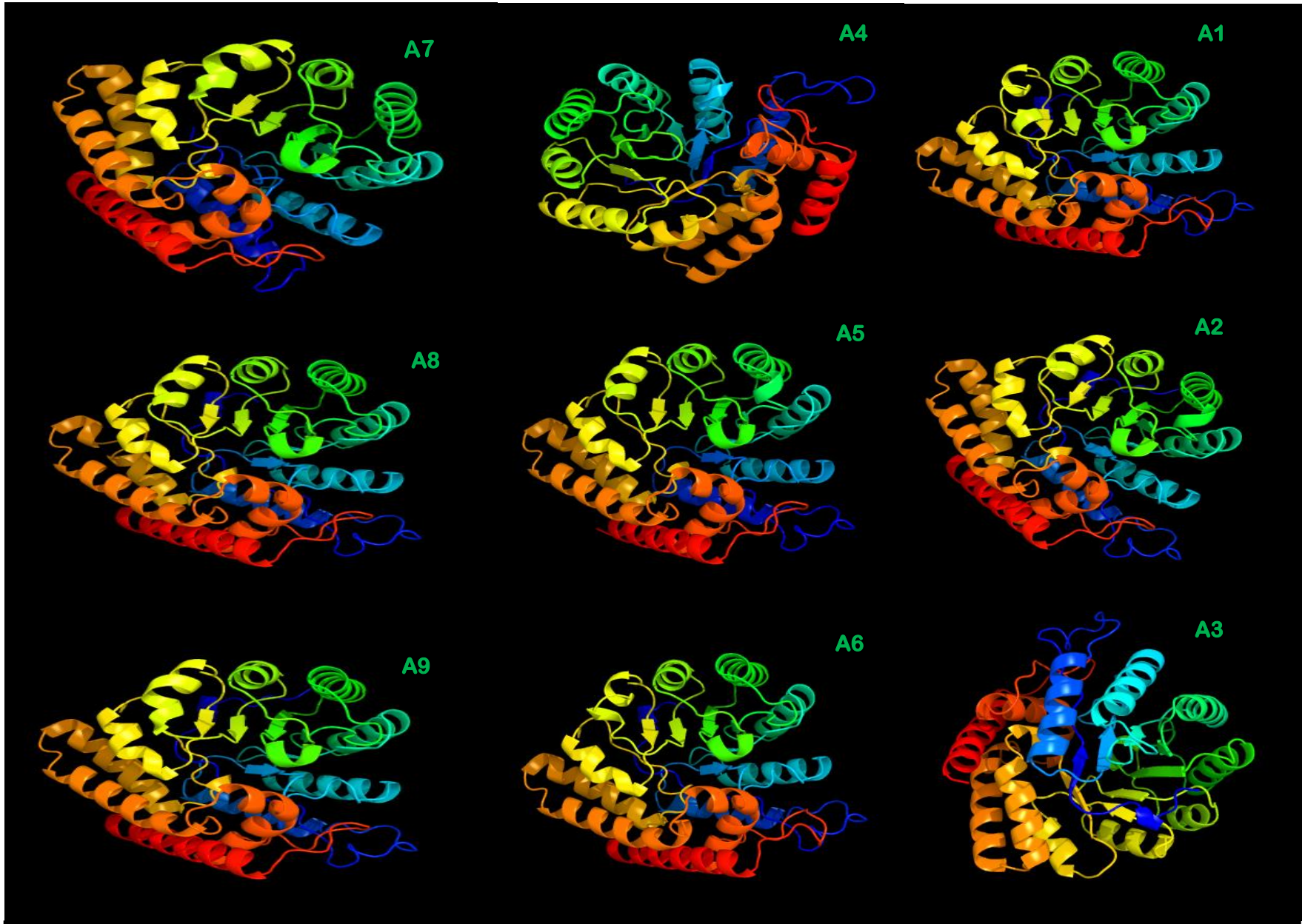

Figure 11: A: 3D structures of pahE Enzymes: 1: Alcaligenes 2. Alteromonas 3. Arthrobacter 4. Burkholderia 5. Citreicella 6: Delftia 7: Polaromonas 8: polycyclovorans 9. Ralstonia

REFERENCES

[1] Liang C, Huang Y, Wang H. pahE, a Functional Marker Gene for Polycyclic Aromatic Hydrocarbon-Degrading Bacteria. Appl Environ Microbiol. 2019 Jan 23;85(3):e02399-18. doi: 10.1128/AEM.02399-18. PMID: 30478232; PMCID: PMC6344622.

[2] Sakshi, Haritash AK. A comprehensive review of metabolic and genomic aspects of PAH-degradation. Arch Microbiol. 2020 Oct;202(8):2033-2058. doi: 10.1007/s00203-020-01929-5. Epub 2020 Jun 6. PMID: 32506150.

[3] Fulekar, M. & Sharma, Jaya. (2008). Bioinformatics applied in bioremediation. Innov. Roma Food Biotechnol. 2.

[4] N. https://www.downtoearth.org.in/news/air-pollution-killing-620000-indians-every-year-global-burden-of-disease-report--40316

[5] A.K. Haritash, C.P. Kaushik,Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review,Journal of Hazardous Materials,Volume 169, Issues 1–3,

[6] Hussar E, Richards S, Lin ZQ, Dixon RP, Johnson KA. Human Health Risk Assessment of 16 Priority Polycyclic Aromatic Hydrocarbons in Soils of Chattanooga, Tennessee, USA. Water Air Soil Pollut. 2012;223(9):5535-5548. doi:10.1007/s11270-012-1265-7

[7] .Niranjan, Vidya & Reddy, Jayarama. (2016). Role of informatics in bioremediation-A biological solution to environmental issues.

[8] Khan, Nida. (2018). Integration of Bioinformatics in Bioremediation. International Journal of Biomedical Data Mining. 07. 10.4172/2090-4924.1000130.

[9] Librando V, Pappalardo M. In silico bioremediation of polycyclic aromatic hydrocarbon: a frontier in environmental chemistry. J Mol Graph Model. 2013 Jul;44:1-8. doi: 10.1016/j.jmgm.2013.04.011. Epub 2013 May 14. PMID: 23732301.

[10] Arun A, Raja PP, Arthi R, Ananthi M, Kumar KS, Eyini M. Polycyclic aromatic hydrocarbons (PAHs) biodegradation by basidiomycetes fungi, Pseudomonas isolate, and their cocultures: comparative in vivo and in silico approach. Appl Biochem Biotechnol. 2008 Dec;151(2-3):132-42. doi: 10.1007/s12010-008-8160-0. Epub 2008 Feb 26. PMID: 18975143.

[11] Librando V, Pappalardo M. Computational study on the interaction of a ring-hydroxylating dioxygenase from Sphingomonas CHY-1 with PAHs. J Mol Graph Model. 2011 Jun;29(7):915-9. doi: 10.1016/j.jmgm.2011.03.001. Epub 2011 Mar 8. PMID: 21441049.

[12] Ghosal, Debajyoti; Ghosh, Shreya; Dutta, Tapan K.; Ahn, Youngho (2016). Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. Frontiers in Microbiology, 7(), –. doi:10.3389/fmicb.2016.01369.

[13] SureshKumar, Pandian, Jibu Thomas, and Vasudevan Poornima. "Structural insights on bioremediation of polycyclic aromatic hydrocarbons using microalgae: a modelling-based computational study." Environmental monitoring and assessment 190.2 (2018): 1-16.

[14] Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995 Dec;11(6):681-4. doi: 10.1093/bioinformatics/11.6.681. PMID: 8808585.

[15] Kelley, L., Mezulis, S., Yates, C. et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053

[16] https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/ramachandran-plot

[17] Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Stereochemical quality of protein structure coordinates. Proteins, 12, 345-364.

[18] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511-9. doi: 10.1002/pro.5560020916. PMID: 8401235; PMCID: PMC2142462.

[19] Messaoudi, Abdelmonaem et al. “Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase.” Theoretical biology & medical modelling vol. 10 22. 2 Apr. 2013, doi:10.1186/1742-4682-10-22

[20] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993 Sep;2(9):1511-9. doi: 10.1002/pro.5560020916. PMID: 8401235; PMCID: PMC2142462.

[21] Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science. 1991 Jul 12;253(5016):164-70. doi: 10.1126/science.1853201. PMID: 1853201.

[22] Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992 Mar 5;356(6364):83-5. doi: 10.1038/356083a0. PMID: 1538787.

[23] Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). J. Appl. Cryst. 26, 283-291.

[24] Teufel, F., Almagro Armenteros, J.J., Johansen, A.R. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol (2022). https://doi.org/10.1038/s41587-021-01156-3

[25] Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol. 2017;1611:59-73. doi: 10.1007/978-1-4939-7015-5_6. PMID: 28451972.

[26] Finn, Robert D et al. “Pfam: the protein families database.” Nucleic acids research vol. 42,Database issue (2014): D222-30. doi:10.1093/nar/gkt1223

[27] Söding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005; 21:951–960.

[28] Jaina Mistry, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A Salazar, Erik L Sonnhammer, Silvio C E Tosatto, Lisanna Paladin, Shriya Raj, Lorna J Richardson, Robert D Finn, Alex Bateman, Pfam: The protein families database in 2021, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D412–D419, https://doi.org/10.1093/nar/gkaa913

[29] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).pp. 571-607

[30] Protein Identification and Analysis Tools on the ExPASy Server;Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).pp. 571-607

[31] Dallakyan, S., & Olson, A. J. (2014). Small-Molecule Library Screening by Docking with PyRx. Chemical Biology, 243–250. doi:10.1007/978-1-4939-2269-7_19

[32] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 931, Naphthalene. Retrieved May 29, 2022 from https://pubchem.ncbi.nlm.nih.gov/compound/Naphthalene.

[33] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 995, Phenanthrene. Retrieved May 29, 2022 from https://pubchem.ncbi.nlm.nih.gov/compound/Phenanthrene.

[34] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 8418, Anthracene. Retrieved May 29, 2022 from https://pubchem.ncbi.nlm.nih.gov/compound/Anthracene.

[35] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 31423, Pyrene. Retrieved May 29, 2022 from https://pubchem.ncbi.nlm.nih.gov/compound/Pyrene.

[36] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 107735, Pyruvate. Retrieved May 29, 2022 from https://pubchem.ncbi.nlm.nih.gov/compound/Pyruvate.

[37] BIOVIA, Dassault Systèmes, [BIOVIA DISCOVERY STUDIO], [CLIENT 2021], San Diego: Dassault Systèmes, [2022].

[38] A.K. Haritash, C.P. Kaushik,Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review,Journal of Hazardous Materials,Volume 169, Issues 1–3,

2009,Pages 1-15,ISSN 0304-3894,https://doi.org/10.1016/j.jhazmat.2009.03.137.

[39] Sanvictores T, Farci F. Biochemistry, Primary Protein Structure. [Updated 2021 Nov 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK564343/

[40] Rehman I, Farooq M, Botelho S. Biochemistry, Secondary Protein Structure. [Updated 2021 Dec 15]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470235/

[41] https://proteinstructures.com/structure/secondary-structure/

[42] https://proteinstructures.com/structure/introduction/

[43] Rehman I, Kerndt CC, Botelho S. Biochemistry, Tertiary Protein Structure. [Updated 2021 Sep 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470269/

[44] https://faculty.uml.edu//vbarsegov/teaching/bioinformatics/lectures/protstructure2modified.pdf

[45] Sun PD, Foster CE, Boyington JC. Overview of protein structural and functional folds. Curr Protoc Protein Sci. 2004;Chapter 17(1):1711-17.1. doi:10.1002/0471140864.ps1701s35

[46] Baker, D. (2001). Protein Structure Prediction and Structural Genomics. Science, 294(5540), 93–96. doi:10.1126/science.1065659

[47] https://en.wikipedia.org/wiki/Structure_validation#Model_validation

[48] Fahim, Arjang et al. “Protein structure validation and identification from unassigned residual dipolar coupling data using 2D-PDPA.” Molecules (Basel, Switzerland) vol. 18,9 10162-88. 22 Aug. 2013, doi:10.3390/molecules180910162

[49] Owji, Hajar; Nezafat, Navid; Negahdaripour, Manica; Hajiebrahimi, Ali; Ghasemi, Youns (2018). A Comprehensive Review of Signal Peptides: Structure, Roles, and Applications. European Journal of Cell Biology, (), S0171933518300189–. doi:10.1016/j.ejcb.2018.06.003

[50] Ingólfsson, Helgi & Yona, Golan. (2008). Protein Domain Prediction. Methods in molecular biology (Clifton, N.J.). 426. 117-43. 10.1007/978-1-60327-058-8_7.

[51] Harnisch U, Weiss H, Sebald W (May 1985). "The primary structure of the iron-sulfur subunit of ubiquinol-cytochrome c reductase from Neurospora, determined by cDNA and gene sequencing". Eur. J. Biochem. 149 (1): 95–9. doi:10.1111/j.1432-1033.1985.tb08898.x. PMID 2986972

[52] Madueño F, Napier JA, Cejudo FJ, Gray JC (October 1992). "Import and processing of the precursor of the Rieske FeS protein of tobacco chloroplasts". Plant Mol. Biol. 20 (2): 289–99. doi:10.1007/BF00014496. PMID 1391772. S2CID 2306978

[53] the primary structure of the iron-sulfur subunit of ubiquinol-cytochrome c reductase from Neurospora, determined by cDNA and gene sequencing. Harnisch U, Weiss H, Sebald W. Eur. J. Biochem. 149, 95-9

[54] Nucleotide sequence and transcription of the fbc operon from Rhodopseudomonas sphaeroides. Evaluation of the deduced amino acid sequences of the FeS protein, cytochrome b and cytochrome c1. Gabellini N, Sebald W. Eur. J. Biochem. 154, 569-79, (1986). View articlePMID: 3004982

[55] The genes of the Paracoccus denitrificans bc1 complex. Nucleotide sequence and homologies between bacterial and mitochondrial subunits. Kurowski B, Ludwig B. J. Biol. Chem. 262, 13805-11, (1987). View articlePMID: 2820981

[56] Import and processing of the precursor of the Rieske FeS protein of tobacco chloroplasts. Madueno F, Napier JA, Cejudo FJ, Gray JC. Plant Mol. Biol. 20, 289-99, (1992). View articlePMID: 1391772

[57] Jakoncic, Jean et al. “The catalytic pocket of the ring-hydroxylating dioxygenase from Sphingomonas CHY-1.” Biochemical and biophysical research communications vol. 352,4 (2007): 861-6. doi:10.1016/j.bbrc.2006.11.117

[58] Nucleotide sequences of the Acinetobacter calcoaceticus benABC genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. Neidle EL, Hartnett C, Ornston LN, Bairoch A, Rekik M, Harayama S. J. Bacteriol. 173, 5385-95, (1991)

[59] Blickling S, Renner C, Laber B, Pohlenz HD, Holak TA, Huber R. Reaction mechanism of Escherichia coli dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. Biochemistry. 1997 Jan 7;36(1):24-33. doi: 10.1021/bi962272d. PMID: 8993314.

[60] Blickling S, Beisel HG, Bozic D, Knäblein J, Laber B, Huber R. Structure of dihydrodipicolinate synthase of Nicotiana sylvestris reveals novel quaternary structure. J Mol Biol. 1997 Dec 12;274(4):608-21. doi: 10.1006/jmbi.1997.1393. PMID: 9417939.

[61] Barbosa JA, Smith BJ, DeGori R, Ooi HC, Marcuccio SM, Campi EM, Jackson WR, Brossmer R, Sommer M, Lawrence MC. Active site modulation in the N-acetylneuraminate lyase sub-family as revealed by the structure of the inhibitor-complexed Haemophilus influenzae enzyme. J Mol Biol. 2000 Oct 27;303(3):405-21. doi: 10.1006/jmbi.2000.4138. PMID: 11031117.

[62] Joerger AC, Mayer S, Fersht AR. Mimicking natural evolution in vitro: an N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. Proc Natl Acad Sci U S A. 2003 May 13;100(10):5694-9. doi: 10.1073/pnas.0531477100. Epub 2003 Apr 23. PMID: 12711733; PMCID: PMC156263.

[63] Meng, Xuan-Yu et al. “Molecular docking: a powerful approach for structure-based drug discovery.” Current computer-aided drug design vol. 7,2 (2011): 146-57. doi:10.2174/157340911795677602

[64] McConkey BJ, Sobolev V, Edelman M. The performance of current methods in ligand-protein docking. Current Science. 2002;83:845–855.

[65] link: https://www.malvernpanalytical.com/en/products/measurement-type/binding-affinity#:~:text=What%20is%20Binding%20Affinity%3F,(e.g.%20drug%20or%20inhibitor).

[66] Pinyakong, Onruthai; Habe, Hiroshi; Omori, Toshio (2003). The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons(PAHs).. The Journal of General and Applied Microbiology, 49(1), 1–19. doi:10.2323/jgam.49.1

[67] Ghosal, Debajyoti; Ghosh, Shreya; Dutta, Tapan K.; Ahn, Youngho (2016). Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. Frontiers in Microbiology, 7(), –. doi:10.3389/fmicb.2016.01369.

[68] Gupta, S., Pathak, B. & Fulekar, M.H. Molecular approaches for biodegradation of polycyclic aromatic hydrocarbon compounds: a review. Rev Environ Sci Biotechnol 14, 241–269 (2015). https://doi.org/10.1007/s11157-014-9353-3

[69] https://www.intechopen.com/chapters/45093

[70] Sobika Bhandari, Darbin Kumar Poudel, Rishab Marahatha, Sonika Dawadi, Karan Khadayat, Sitaram Phuyal, Shreesti Shrestha, Santosh Gaire, Kusum Basnet, Uddhav Khadka, Niranjan Parajuli, "Microbial Enzymes Used in Bioremediation", Journal of Chemistry, vol. 2021, Article ID 8849512, 17 pages, 2021. https://doi.org/10.1155/2021/8849512

[71] Yen KM, Gunsalus IC. Regulation of naphthalene catabolic genes of plasmid NAH7. J Bacteriol. 1985 Jun;162(3):1008-13. doi: 10.1128/jb.162.3.1008-1013.1985. PMID: 3997772; PMCID: PMC215876.