

Discovery of Highly Conserved Evolutionary Regions in Polyhydroxyalkanoate Synthase (PhaC) From *Pseudomonas* sp. Using Bioinformatics and *In-silico* Approaches

Ubaid Mehmood^{1,2‡}, Faizan Muneer^{1‡}, Xiao Junhua², Habibullah Nadeem^{*1}

¹Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Faisalabad, Pakistan.

²College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, 2999 Renmin North Road, Shanghai, 201620, P.R. China

‡ Authors with equal Contributions

*For Correspondence: Dr. Habibullah Nadeem (habibullah@gcuf.edu.pk)

DOI: 10.29322/IJSRP.12.07.2022.p12702

<http://dx.doi.org/10.29322/IJSRP.12.07.2022.p12702>

Paper Received Date: 10th June 2022

Paper Acceptance Date: 27th June 2022

Paper Publication Date: 6th July 2022

Abstract- Microbial polyesters such as polyhydroxyalkanoates (PHAs) are biopolymers with unique physicochemical and thermal properties that can replace synthetic polymers from the environment. The ability of these polymers to aid multiple processes in pharmaceuticals, drug delivery systems, and bioplastic materials has increased their worth in the polymer industry. Polyhydroxyalkanoate synthase (PhaC) has a crucial role in the biosynthesis of PHA and is produced by various microbes. *Pseudomonas* strains have a high potential to produce PhaC. Due to limited structural, functional, and phylogenetic studies of the PhaC, its important evolutionary and underlying molecular functional characteristics remain unknown. This study envisaged finding out the phylogenetic relation, conserved motifs, and 3D structural coherence in PhaC for various species of the *Pseudomonas*. Three highly conserved evolutionary regions named R1, R2, and R3 were identified in forty strains of *Pseudomonas* including five conserved regions in each phylogenetic group formed. R1 with amino acid sequence of (Arg-263)-(Glu-264)-(Trp-265)-(Gly-266)-(Leu-267), R2 with (Phe-367)-(Ala-368)-(Trp-369)-(Met-370)-(Arg-371)-(Pro-372)-(Asn-373)-(Asp-374)-(Leu-375)-(Ile-376) and R3 having (Asp-452)-(His-453)-(Ile-454)-(Thr-455)-(Pro-456)-(Trp-457) amino acid sequence were determined. Ramachandran plot analysis for the PhaC of *P. fluorescens*, *P. stutzeri*, *P. putida*, *P. aeruginosa*, and *P. syringae* suggested that nearly 87% of the residues of PhaC are in favored regions with 11-12% in additionally allowed and 1-2% in allowed regions. This study has confirmed for the first time the presence of three highly conserved in the PhaC protein of the *Pseudomonas* genus additionally with conserved sites in each phylogenetic group that can help in understanding and improving the function of this protein in the future.

Keywords - Bioplastics; Carbonosomes; Evolution; Polyesters; Phylogenetic; Ramachandran plot

I. INTRODUCTION

Polyhydroxyalkanoates (PHAs) are polyesters produced by a wide range of microbes under nutrient stress and unfavorable environmental conditions [1]. These polyesters are also termed Carbonosomes by the researchers working in the polymeric fields because they are produced in the presence of excess carbon sources. Prokaryotes such as bacteria and some microalgae and fungi produce PHA as cytoplasmic inclusions of 0.2-0.5 μm [2]. Studies have confirmed that these microbial inclusions are energy reserves and protect the bacterial cells from osmotic stress and environmental imbalances [3]. PHAs are diverse and versatile due to the difference in monomeric compositions and the number of carbon atoms in the monomer unit of the polyesters [4]. Maurice Lemoigne, a French chemist, first isolated these polyesters from *Bacillus megaterium* in the

1920s and showed that these bacterial polymers can be used to form transparent films [5]. Depending upon the number of monomers in the polyester chain, PHAs are usually divided into three major groups namely, short-chain length PHA (scl-PHA), medium-chain length PHA (MCL-PHA), and long-chain length PHA (LCL-PHA) [6]. Scl-PHA have 3-4 carbon atoms and are the most common forms. Poly-3-hydroxybutyrate (P3HB) and poly-4-hydroxybutyrate (P4HB) are common examples of scl-PHA. Mcl-PHA such as poly-3-hydroxyhexanoate (P3HHx), and poly-3-hydroxyoctanoate (P3HO) have 6-14 carbon atoms. Mcl-PHA can be either homopolymer such as (P3HHx) and (P3HO) or copolymers like poly-3-hydroxybutyrate-co-3-hydroxyvalerate (P(3HB-co-3HV)) [2]. Long-chain PHAs (lcl-PHA) are highly complex and have not been considered good polymers for bioplastic production [2].

PHA is biopolymers that are not only ecofriendly and sustainable but important bio-based raw material, used in bioplastic production [7]. Various bacterial species such as *Bacillus megaterium*, *Cupriavidus nector*, *Ralstonia eutropha*, and *Pseudomonas* sp. have a high potential to produce PHA. *Pseudomonas* species usually produce mcl-PHA and were studied for the first time in 1983 [2]. Since the discovery of the first PHA in the polymeric form of P3HB between 1923 and 1927 by Lemoigne, more than 300 microbial species including bacteria, algae, and fungi have been identified to produce these eco-friendly and sustainable polyesters [2]. PHA shows interesting polymeric properties such as glass transition temperature (T_g), melting temperature (T_m), crystallinity, tensile strength, extension to break ratios, and Young's modulus [8]. Excellent physicochemical and mechanical properties make PHA a natural, eco-friendly, and sustainable replacement for synthetic or petrochemical-based polymers (plastics) such as polyethylene (PE), polypropylene (PP), and polyethylene terephthalate (PET), and polyvinyl chloride (PVC) [2]. Synthetic plastics since their first prominent discovery back in 1907 have ruled the world's polymeric industry and research [9]. Plastics got much focus on the development of industrial and domestic products that their presence is undeniable today. From small-scale electronic appliances to large-scale heavy machinery, all employ some sort of plastic materials in their production [10].

Daily life products, pharmaceuticals, packaging materials, and cosmetics are dependent on synthetic plastics one way or the other during their production line due to the lack of natural, bio-based polymers that might be an efficient replacement [11]. Polyhydroxyalkanoates as biopolymers have thermo-mechanical properties similar to synthetic plastics therefore, these are the potential candidates for the development of bioplastics. The sharp and continuously growing production curve of synthetic plastics is alarming for the global environment, health, and economy [8]. Climate change, environmental pollution, destruction of ecosystems, biosphere, and public health are currently in their worse scenarios due to the severe impacts of plastics [12]. Today, almost 340-350 million tons of synthetic plastics are produced globally [2]. Limited recycling processes, production, and release of toxic gases during the incineration of plastics, less resource for plastic waste management are the major factors that have proved to a stimulus towards the need for bio-based plastics referred to as bioplastics [13].

Bioplastics are made using various biomass sources including agro-polymers (starch, cellulose), animal proteins (keratin, chitin, chitosan), and microbial polyesters (PHA) [14][15]. Although a huge amount of biomass can be obtained from agro-polymers and animal proteins for the production of biomaterials, microbial polyesters are considered more versatile, sustainable, and ecofriendly [16]. Polyhydroxyalkanoates (PHAs) are, therefore, potential candidates for the development of bioplastics that can help in achieving the sustainable development goals (SDGs) of the United Nations by countering and replacing petroleum-based synthetic plastics [17]. One thing that makes PHA a more crucial candidate for bioplastic and biomaterial development is the variety in its polymeric forms obtained under varying synthesis environments coupled with a wide range of nutrients and bacterial strains. Researchers can, therefore, use specific microorganisms and provide them controlled environment to obtain polymers of their own choice and physicochemical characteristics [18]. *Azotobacter vinelandii*, *Bacillus megaterium*, and *Bacillus licheniformis* produce scl-PHA or more specifically P3HB under glucose, lactose as carbon sources, and incubation temperature between 37–40 °C. *Pseudomonas putida* KT2440, *P. chlororaphis*, and *P. mendocina* biosynthesized mcl-PHA when providing various carbon sources and an incubation temperature of 37–45 °C for 72 hours [2]. A study of the PhaC concerning its structure, functions, catalytic sites, and evolution is highly desirable to design and produce PhaC with highly improved and efficient properties for hyper-production of PHA [19].

This study focused on the structural and functional analysis of PHA synthase protein (PhaC) using different bioinformatics and *in silico* approaches. Structural analysis using bioinformatics tools and computational methods can help in the indirect evidence of PhaC and its varying response to different carbon substrates during PHA synthesis and its catalytic activity [20]. The study of catalytic residues in the active site of PhaC for different bacterial species and their phylogenetic analysis on the same bases can help biotechnologists to create recombinant microbes and PHA synthases for the hyper-production of PHA making the process not only highly efficient but also economically feasible.

II. MATERIAL AND METHODS

PROTEIN SEQUENCE RETRIEVAL FOR PHA-PRODUCING *PSEUDOMONAS* STRAINS

Forty different strains of *Pseudomonas* were selected from the available literature based on their ability to produce PHA. Polyhydroxyalkanoate synthase (PhaC) protein sequences for the selected forty strains were retrieved from uniprot (<https://www.uniprot.org/>) and NCBI (www.ncbi.nlm.nih.gov) in FASTA format and saved in word file for further analysis [21].

CONSTRUCTION OF PHYLOGENETIC TREE

Based on the PHA synthase (PhaC) protein sequence retrieved from the NCBI database a phylogenetic tree was constructed using MEGA7 software. ClustalW was used for aligning the amino acid sequence of the protein and BLASTp the sequence to search for the similarity and common residue in the sequences [22]. Results were used to construct a maximum likelihood phylogenetic tree. The tree was further divided into different groups (A, B, and C, based on the similarity and conserved motifs of the proteins among different strains of *Pseudomonas*. One strain from each group was designated as the representative for the whole group by hypothesizing that each group has identical and conserved residues. *P. fluorescens*, *P. stutzeri*, *P. putida*, *P. aeruginosa* PA01, and *P. syringae* were used as representatives from group A respectively.

PREDICTION OF SECONDARY AND 3D STRUCTURES OF SELECTED PhaC PROTEIN

Chou and Fasman's secondary structure prediction server (CFSSP) (<http://www.biogem.org/tool/chou-fasman/index.php>) [23] was used to predict the unknown secondary structure of the PhaC proteins from the selected *Pseudomonas* strains. The possible 3D structures were predicted using Modeller 9.20 software. PyMOL software was used for analyzing the 3D ribbon structure of PhaC using PDB files of the amino acid sequences for the selected strains. The strains and their protein IDs were *P. fluorescens* (ACK57562.1), *P. stutzeri* (AAO59384.1), *P. putida* (AAP37050.1), *P. aeruginosa* PA01 (NP_253745.1) and *P. syringae* (WP_024673311.1).

PHYSIOCHEMICAL CHARACTERIZATION OF PHAC FOR THE SELECTED STRAINS OF *PSEUDOMONAS*

Physiochemical characterization of PhaC from various *Pseudomonas* strains was accomplished by using Protparam (<https://web.expasy.org/protparam/>) [24].

RETRIEVAL AND 3D STRUCTURAL ANALYSIS OF PHAC OF *CHROMOBACTERIUM* SP. AND *RALSTONIA EUTROPHA* FOR COMPARISON WITH THE PREDICTED STRUCTURES OF PhaC

The already predicted and confirmed structures of PhaC protein for *Chromobacterium* sp. USM2, (PDB ID 5XAV and 6K3C) and *Ralstonia eutropha* (PDB ID 5HZ2) were retrieved from the protein data bank (www.rcsb.org). Retrieved PDB files were analyzed using PyMOL [25]. Active site and residues actively participating in binding at the site were determined. The results were compared to that of the predicted structures of *Pseudomonas* strains PhaC protein to find the evolutionarily conserved motifs of PhaC proteins in different bacterial strains and to determine the residual chains of active sites of the protein [2].

III. RESULTS AND DISCUSSION

PROTEIN SEQUENCE RETRIEVAL FOR PHA-PRODUCING *PSEUDOMONAS* STRAINS

Sequence retrieval for PhaC protein of various *Pseudomonas* strains from NCBI showed that almost all the strains have PhaC protein consisting of 560 amino acids and a mass of 62544 (62.544 kDa) to 62809 Da (62.809 kDa). Due to the equal number of amino acids and molecular mass, the enzyme is almost similar in its properties dependent on mass such as the isoelectric point of the protein [24].

PHYLOGENETIC ANALYSIS

Retrieved PhaC protein sequences from the Uniprot database were used to construct a maximum likelihood tree using Mega7 (Fig. 1). The tree is divided into 5 groups depending upon the sequence identity in the conserved motifs. Group-A is shown in blue color and has 17 strains of *Pseudomonas* with *P. synxantha* at the top and *P. cichorii* at the bottom of the group. The yellow color represented group B which has three strains and is the second smallest group of the phylogenetic tree with *P. stutzeri* being at the top and *P. mendocina* at the bottom. Group-C has six bacterial strains and is shown in green. *P.*

parafulva and *P. monteilii* occupy the top and bottom of the group respectively. The smallest group of the phylogenetic tree is group-D and has only two strains namely *P. aeruginosa* PA01 and *P. denitrificans*, which are colored purple. Group-E indicated with red color and has 12 strains. *P. fulva* and *P. avellanae* are present at the top and bottom of this group respectively. For each group, the PhaC sequence was aligned separately to find out the conserved residues or more specifically the conserved motifs. Alignment results for each group can be seen in Fig 2a, 2b, 2c, 2d, and 2e.

Group-A has a conserved motif sequence of 28 amino acids for the PhaC protein that is composed of *Val-Asn-Leu-Gly-Ala-Cys-Ala-Gly-Gly-Thr-Ile-Ala-Ala-Leu-Gln-Gly-His-Leu-Gln-Ala-Lys-Arg-Gln-Arg-Arg*. The asterisk mark in the chain showed a variable amino acid. In the *P. chlororaphis* PhaC motif, the sequence has a methionine (Met) replaced with leucine (Leu) at the first variable site. At the second variable amino acid site in the motif sequence, *P. synxantha*, *P. aurantiaca*, and *P. corrugata* have threonine (Thr), methionine (Met), and threonine (Thr) respectively, while all other strains have leucine (Leu) at this position. The third and last variable site in the conserved motif of group-1 is for *P. breunneri*, which has glutamine (Gln) instead of leucine (Leu) at this position, unlike other strains. Group-B has a conserved motif of 28 amino acids for PhaC that account for the amino acid chain of *Arg-Pro-Leu-Leu-Ile-Val-Pro-Pro-Gln-Ile-Asn-Lys-Tyr-Tyr-Ile-Phe-Asp-Ser-Leu-Asn-Asp-Lys-Ser-Phe-Val-Gln-Tyr-Ala*. The motif sequence for group-B is almost identical with no variable amino acid site in the sequence.

Group-C has 27 amino acids in the conserved region with one variable position of amino acid in the group. The amino acid sequence with high similarity is composed of *Asp-Ser-Phe-Thr-Val-Ala-Gly-Ser-Asn-Asp-His-Ile-Thr-Pro-Trp-Asp-Ala-Val-Tyr-Arg-Ser-Ala-Leu-Leu-Gly-Gly*. In *P. parafulva* the variable amino acid site has a threonine (Thr) unlike all other members of this group that have leucine (Leu). *P. aeruginosa* PA01 and *P. denitrificans* are the only members of group D and have a conserved motif of 49 amino acids namely *Met-Arg-Glu-Lys-Gln-Glu-Ser-Gly-Ser-Val-Pro-Val-Pro-Ala-Glu-Phe-Met-Ser-Ala-Gln-Ser-Ala-Ile-Val-Gly-Leu-Arg-Gly-Lys-Asp-Leu-Leu-Thr-Thr-Val-Arg-Ser-Leu-Ala-Val-His-Gly-Leu-Arg-Gln-Pro-Leu-His-Ser*. PhaC sequence of group-D is identical for both strains with no variable site. Group-E of the *Pseudomonas* strain has 12 members with a conserved motif of 45 amino acids that include *Ser-Tyr-Gln-Ala-Gly-Val-Leu-Gly-Gly-Asp-Met-Ala-Lys-Val-Phe-Ala-Trp-Met-Arg-Pro-Asn-Asp-Leu-Ile-Trp-Asn-Tyr-Trp-Val-Asn-Asn-Tyr-Leu-Leu-Gly-Asn-Glu-Pro-Pro-Phe-Asp-Ile-Leu*. The sequence has two variable sites with the first variation in *P. fulva* and *P. viridiflava* where arginine (Arg) is replaced with lysine (Lys) and serine (Ser) respectively.

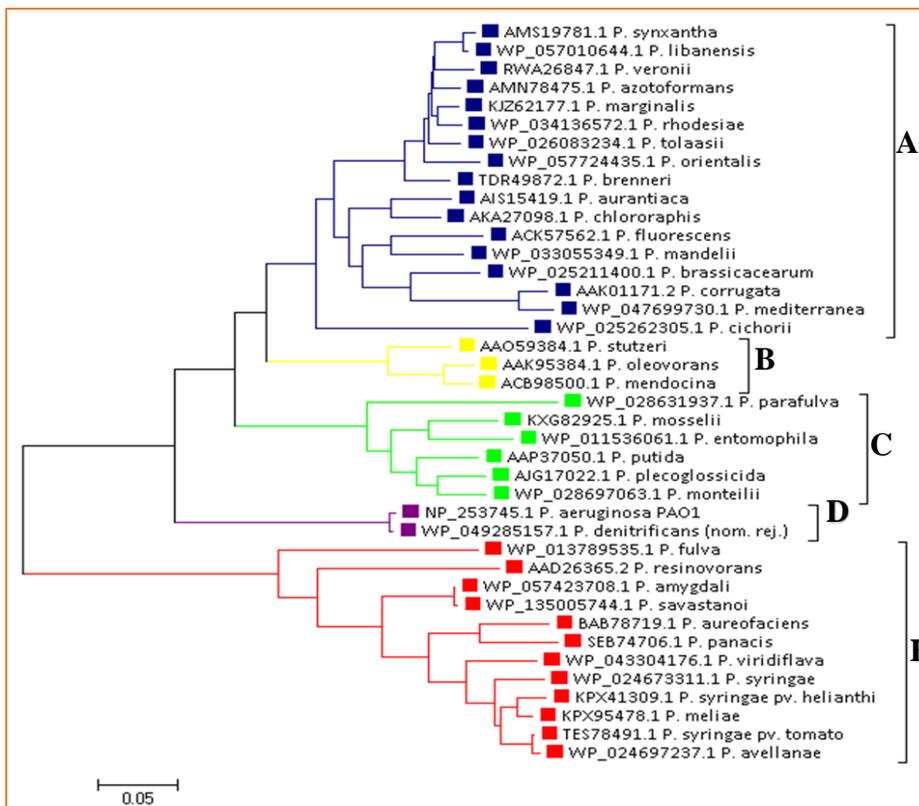


Fig. 1. Maximum likelihood tree constructed for forty different strains of *Pseudomonas*. PhaC protein sequence was used to infer the tree in MEGA7 software. The tree has five groups that are named Group-A to E depending on the evolutionarily conserved motif sequences in the PhaC protein. Members of each group are evolutionary more conserved and similar in terms of their PhaC protein as compared to members of any other group. A total of forty *Pseudomonas* strains are present in the tree with each group having a variable number of *Pseudomonas* strains in it. Group-A has 17 members and is highlighted as a blue taxon. Group B is shown yellow and has three members. Green taxon represented the six members of Group-C. Group D is the smallest group of the tree and is shown with purple color. Group-E has 12 members and is shown in red.

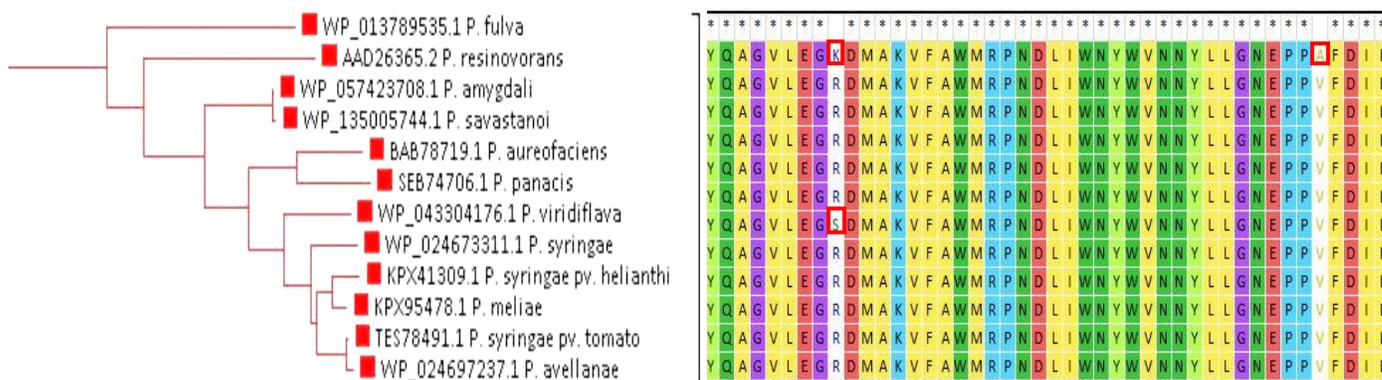
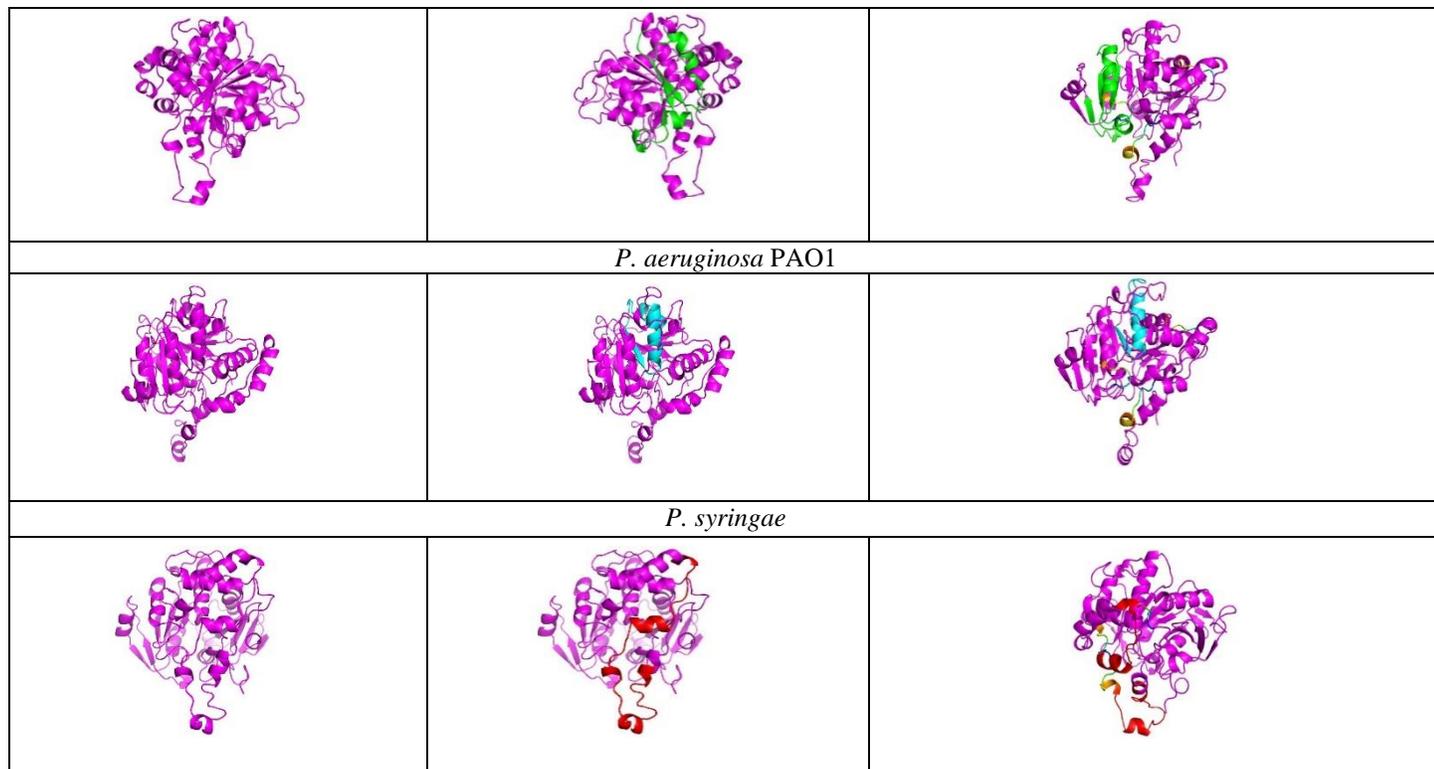


Fig. 2e. Bacterial strains in group-E have 45 amino acids in the conserved PhaC motif sequence with three variable sites. *P. fulva* has two variable amino acid sites. Lysine (Lys) and alanine (Ala) replace arginine (Arg) and valine (Val) respectively unlike other strains except for *P. viridiflava* where arginine (Arg) is replaced with serine (Ser).

Based on identical residues of each group 3D protein structures were drawn in PyMOL that help us to understand how evolution has conserved specific residual chains in these groups. A comparison of the predicted 3D structure along with conserved sites in groups and the selected strains of *Pseudomonas* from each group are represented in Table 1.

Table 1. A comparison of common residual chains of the PhaC protein of *Pseudomonas*.

3D-structure of PhaC	Common residual chains among the phylogenetic group	Common residual chains in selected <i>Pseudomonas</i> species
<i>P. fluorescens</i>		
<i>P. stutzeri</i>		
<i>P. putida</i>		



SECONDARY AND 3-D STRUCTURES OF PhaC PROTEIN

Simple secondary structures, as well as 3D structures of PhaC protein for five different strains of *Pseudomonas*, were obtained. Each strain was selected to represent the whole group because all other *Pseudomonas* sp. in a group has identical sequence and conserved motifs. Chou and Fasman's secondary structure prediction server (CFSSP) (<http://www.biogem.org/tool/chou-fasman/index.php>) [23] was used to predict the unknown secondary structure of the selected *Pseudomonas* strains. Fig. 3a – e. secondary structure prediction showed that all the selected strains have almost identical secondary structures for PhaC protein with minor differences in the percentage of α -helices, β - sheets, and turns. *P. fluorescens* PhaC has 63.4 % α -helices, 65.2 % β -sheets and 15.2 % turns. In the secondary structure of PhaC of *Pseudomonas stutzeri* 67.9 % α -helices, 63 % β - sheets, and 15.5 % turns were present. In *Pseudomonas putida* secondary structure was composed of 57.3 % α -helices, 59.1 % β - sheets, and 15 % turns. 69.8 % α -helices, 64.5 % β - sheets, and 15.7% turns were predicted in the secondary structure of *Pseudomonas aeruginosa* PAO1 PhaC. *Pseudomonas syringae* had 73.9% α -helices, 61.5% β - sheets and 14.1 % turns in the secondary structure of PhaC protein. The data showed that PhaC of *P. syringae* has the highest percentage of α -helices i.e. 73.9 % . *P. fluorescens* had 65.2% β - sheets, which was the highest among the selected strains. *Pseudomonas aeruginosa* PAO1 had the maximum percentage i.e. 15.7 % of turns in the secondary structure of PhaC. The 3D structure of the selected *Pseudomonas* strains predicted using Modeller 9.20 software can be seen in Fig. 4.

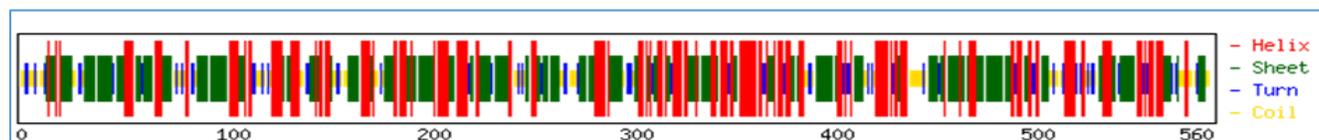


Fig. 3a. Secondary structure of PhaC protein of *Pseudomonas fluorescens* highlighting the α -helices, β - sheets, and turns.

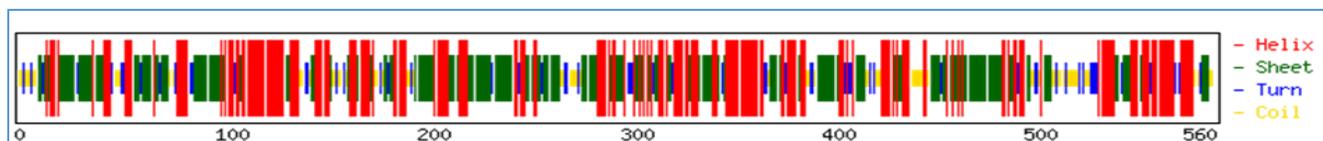


Fig. 3b. Secondary structure of PhaC protein of *Pseudomonas stutzeri* highlighting the α -helices, β - sheets, and turns.

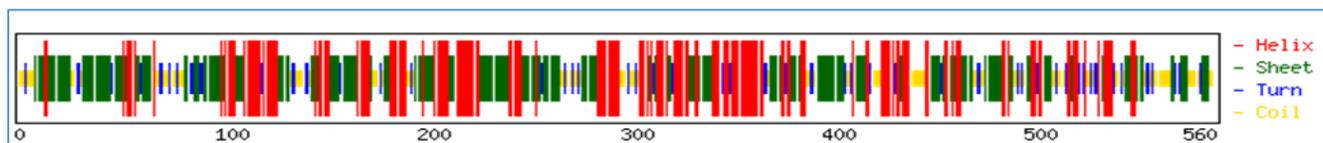


Fig. 3c. Secondary structure of PhaC protein of *Pseudomonas putida* highlighting the α -helices, β - sheets, and turns.

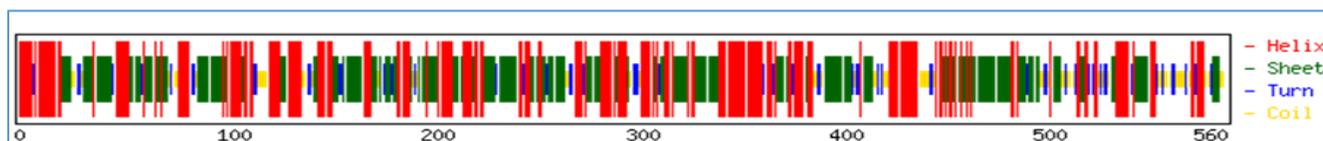


Fig. 3d. Secondary structure of PhaC protein of *Pseudomonas aeruginosa* PAO1 highlighting the α -helices, β - sheets, and turns.

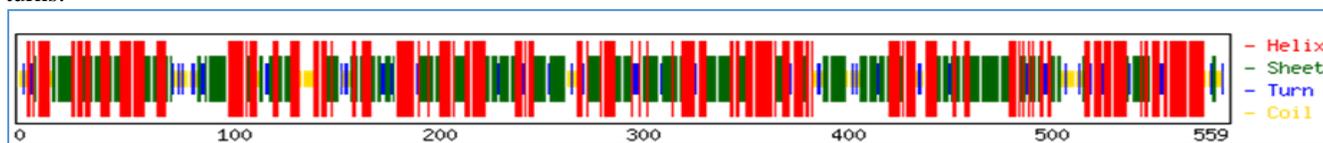


Fig. 3e. Secondary structure of PhaC protein of *Pseudomonas syringae* highlighting the α -helices, β - sheets, and turns.

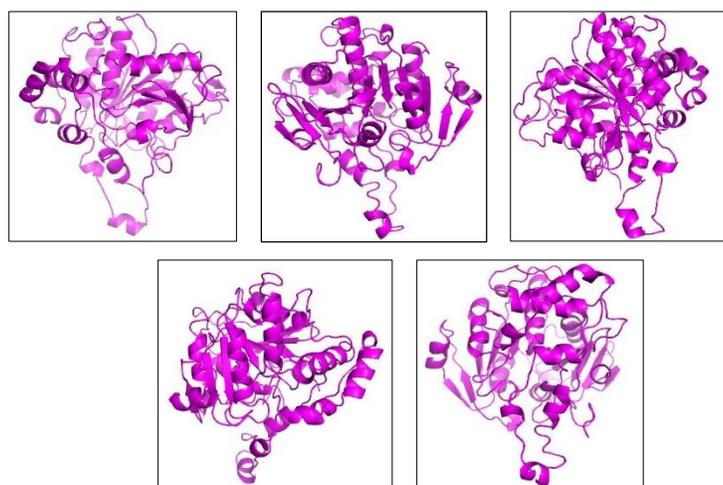


Fig. 4 (a-e). Predicted 3D structures of PhaC for *Pseudomonas fluorescens*, *Pseudomonas stutzeri*, *Pseudomonas putida*, *Pseudomonas aeruginosa* PAO1 and *Pseudomonas syringae* respectively.

PHYSIOCHEMICAL CHARACTERIZATION OF PhaC

The physiological characteristics of the PhaC proteins are observed by Protparam online tool [21]. The huge majority estimation in this server indicated the protein stability and steadiness, in the light of the fact that the steadiness and stability were identified with its appropriate functional ability shown in Table 2. The data represented that higher instability index

value (>40) for PhaC protein makes it unstable. Negative GRAVY values indicated that the protein is non-polar or hydrophilic, which makes it easier to be resolved on a 2-D gel.

Table. 2. PhaC Protparam characterization enlisting important physiochemical properties.

Organism	Amino acids	Molecular mass (Da)	Theoretical isoelectric point (pI)	Ext. coefficient	Instability index	Aliphatic index	GRAVY
<i>P. fluorescens</i>	560	62805.42	9.16	103945	44.46	86.57	-0.324
<i>P. stutzeri</i>	560	62648.31	9.20	109570	45.14	89.95	-0.315
<i>P. putida</i>	560	62809.42	9.35	109445	45.65	86.27	-0.402
<i>P. aeruginosa</i> PAO1	560	62738.41	7.79	109445	50.17	94.07	-0.253
<i>P. syringae</i>	559	62544.28	6.77	102245	43.13	85.87	-0.291

COMPARATIVE ANALYSIS OF PhaC OF *CHROMOBACTERIUM* SP. AND *RALSTONIA EUTROPHA* WITH THE PREDICTED STRUCTURES OF PhaC FOR *PSEUDOMONAS* STRAINS WITH THE DETERMINATION OF ACTIVE AND BINDING SITES

Comparative analysis of PhaC 3-D structures for *Chromobacterium* sp. USM2 and *Ralstonia eutropha* with the predicted structures of PhaC of *Pseudomonas* strains was performed. It showed that the binding ability of the catalytic site of *Pseudomonas* strains is different as compared to that already predicted in *Chromobacterium* sp. USM2 (PDB ID 6K3C), *Ralstonia eutropha* (PDB ID 5HZ2), and *Pseudomonas* sp. UMS 4-55 by Wahab et al., (2006). The 3D protein structures with the catalytic sites of the *Chromobacterium* sp. USM2 and *Ralstonia eutropha* can be seen in Fig. 5 and 6 respectively.

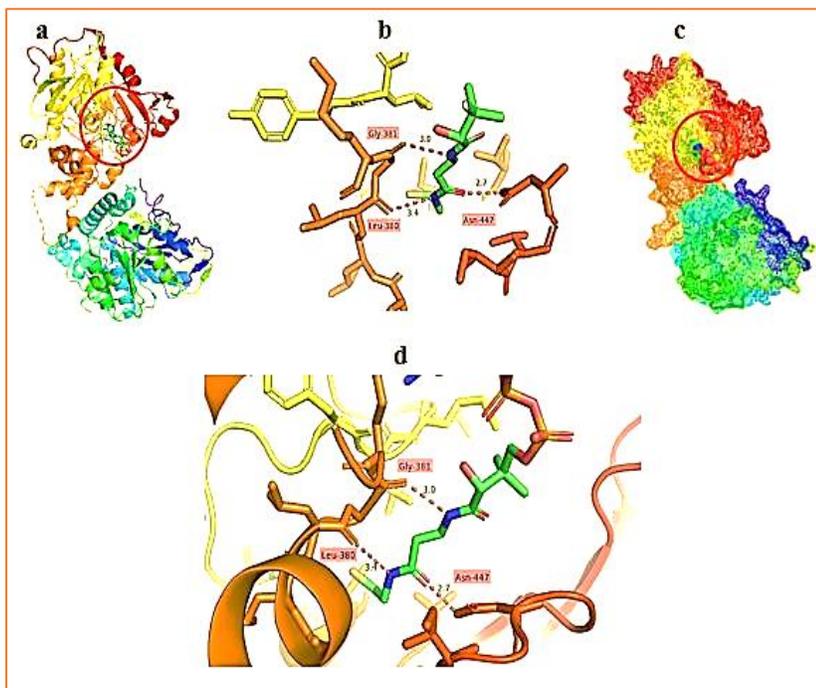


Fig. 5. PhaC of *Chromobacterium* sp. USM2 (PDB ID 6K3C) was analyzed using PyMOL. (a) 3D ribbon structure with CoA (Coenzyme-A) residue bound at the active site (encircled red) (b) Active site of PhaC of *Chromobacterium* (c) Mesh structure of PhaC with CoA shown as balls at the active site (encircled red) (d) Active site binding CoA residue with three active residues (Leucine-380, Glycine-381, and Asparagine-447) at the active site.

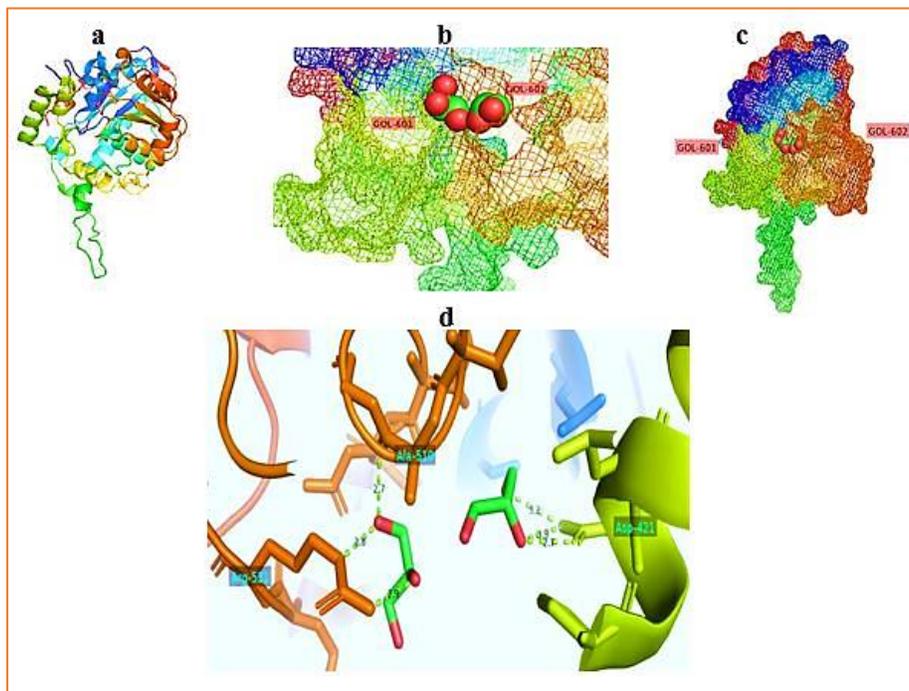


Fig. 6. PhaC of *Ralstonia eutropha* (PDB ID 5HZ2) analyzed through PyMOL (a) 3D ribbon structure of PhaC with two glycerol molecules as substrate bound at the active site (encircled red). (b) Mesh structure of the active site with glycerol molecules (GOL-601 and GOL-602) encircled red. (c) Mesh structure of PhaC with residues at the active site. (d) Residues GOL-601 and GOL-602 bound at the active site of PhaC of *R. eutropha* with the amino acids asparagine-421 (Asp-421), alanine-510 (Ala-510) and arginine-521 (Arg-521).

According to the results, the catalytic sites of these two strains are although identical when compared in terms of their residues but the amino acids that take an active part in binding the substrate are different. *Chromobacterium* sp. USM2 has leucine-380, glycine-381, and asparagine-447 at the active site while asparagine-421, alanine-510, and arginine-521 are present in the active site of *Ralstonia eutropha*. Searching for the identical sequence within the members of each group and the complete phylogenetic tree we concluded that; (1) all the strains given in the phylogenetic tree have three highly conserved regions named R1, R2, and R3. R1 has the amino acid sequence of Arg-263-Glu-264-Trp-265-Gly-266-Leu-267. R2 has the amino acid sequence of Phe-367-Ala-368-Trp-369-Met-370-Arg-371-Pro-372-Asn-373-Asp-374-Leu-375-Ile-376 and R3 having Asp-452-His-453-Ile-454-Thr-455-Pro-456-Trp-457 sequence (Fig. 7-11) for the conserved sites and the Ramachandran plots for PhaC proteins for each selected *Pseudomonas* strain [26][27]. Sequences that are conserved in bacterial groups (group A to E) can also be seen in Fig. 7–11. It was found that while the regions remain intact for their locations and sequence the conserved sites were variable among groups. This suggested that while evolution has conserved important functional residues in all the *Pseudomonads* species different habitats and natural selection have conserved different sequences in various groups of the bacteria. The 3D structures of the conserved sites R1, R2, and R3 are shown in Fig. 12 a-c while the 3D structures of the conserved sites among the selected groups of *Pseudomonas* can be seen in Fig. 13 a-e.

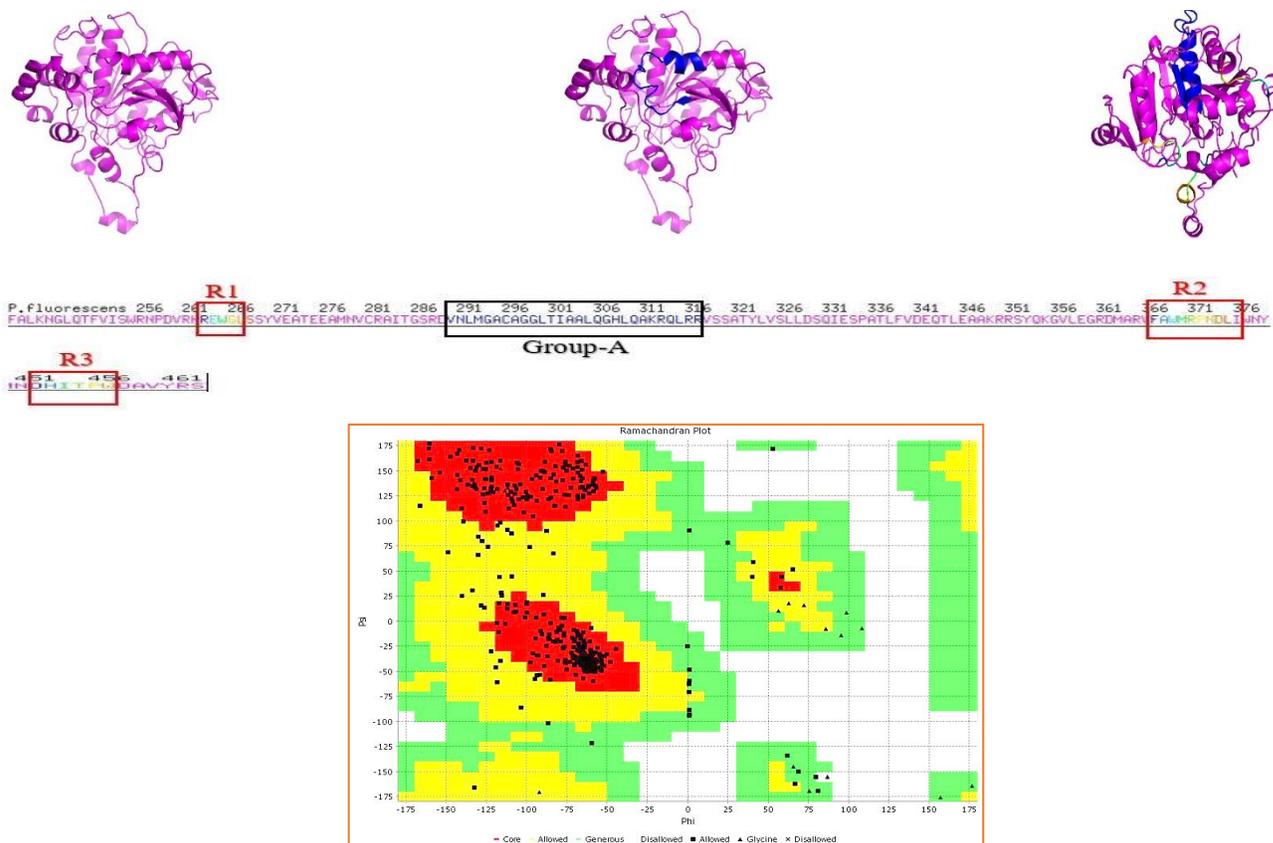
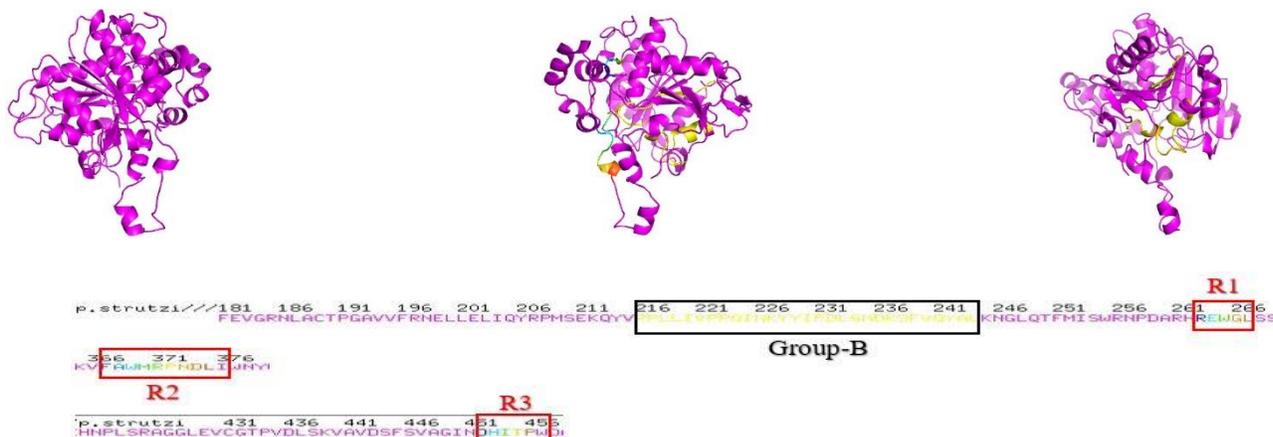


Fig.7. Evolutionary conserved regions R1, R2, and R3 of PhaC in *Pseudomonas* strain with the group-A conserved site for *P. fluorescens*. Ramachandran plot for the protein showed that 86.8% of the residues are in favored regions. The graph further indicated that almost 11.1% of amino acids are additionally allowed and 2.1 % in generously allowed regions. No residue of the protein was present in the disallowed region.



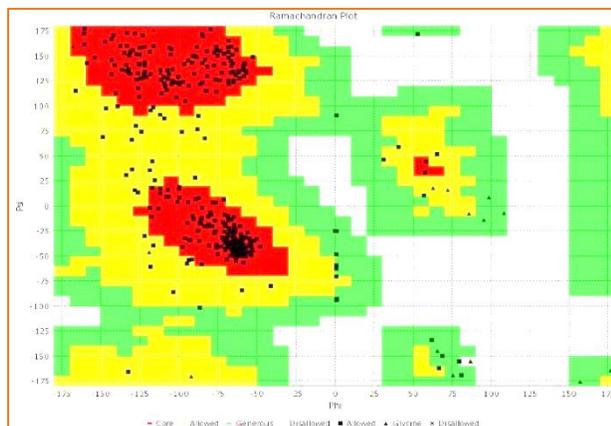


Fig. 8. Evolutionary conserved regions R1, R2, and R3 of PhaC in *Pseudomonas* strain with the group-B conserved site for *P. stutzeri*. Ramachandran plot for the protein showed that 87.3% of the residues were in favored regions. The graph further indicated that almost 11.2% of amino acids were in additionally allowed and 1.5 % in generously allowed regions. No residue of the protein was present in the disallowed region.

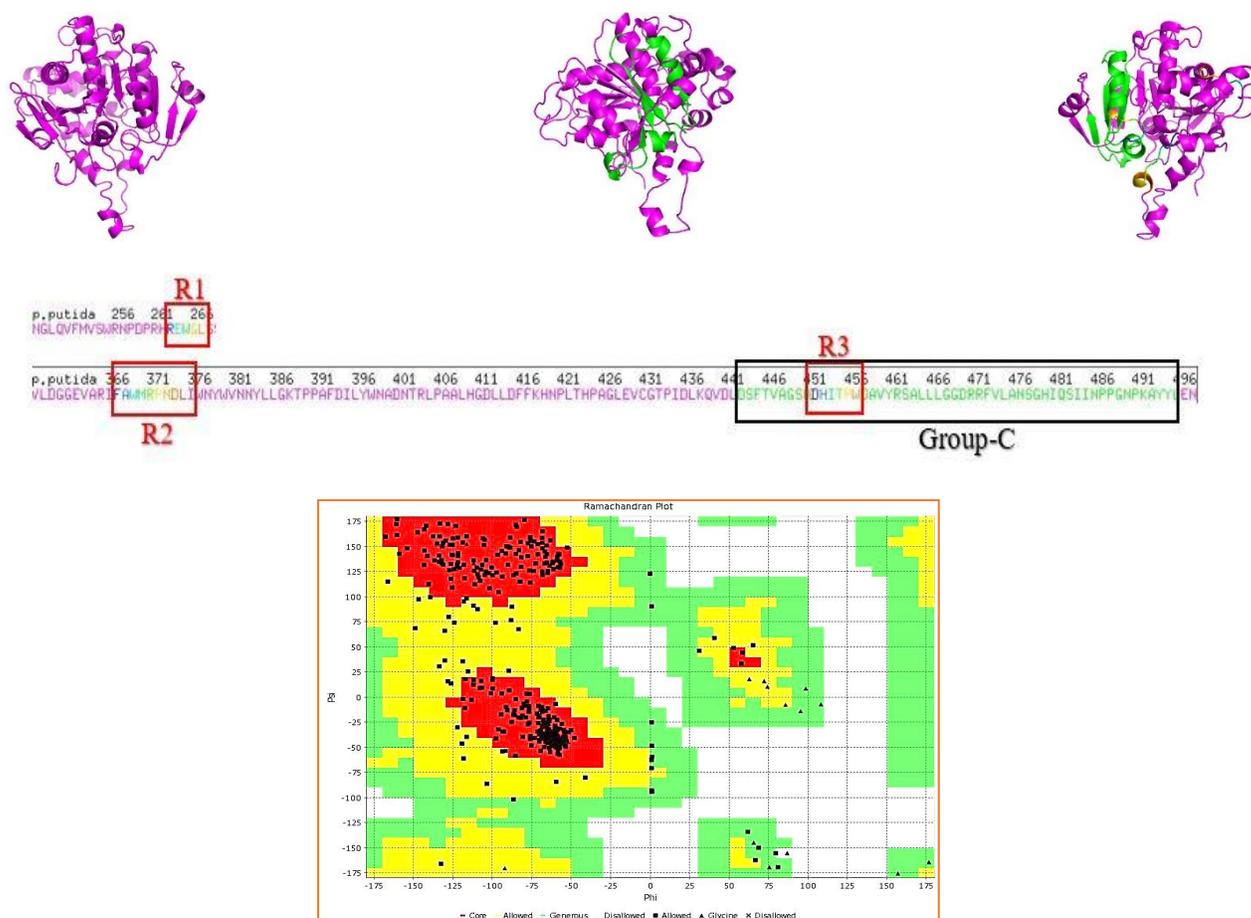


Fig. 9. Evolutionary conserved regions R1, R2, and R3 of PhaC in *Pseudomonas* strain with the group-C conserved site for *P. putida*. Ramachandran plot for the protein showed that 86.6% of the residues were in favored regions. The graph further

indicated that almost 11.8% of amino acids were in additionally allowed and 1.6 % in generously allowed regions. No residue of the protein was present in the disallowed region.

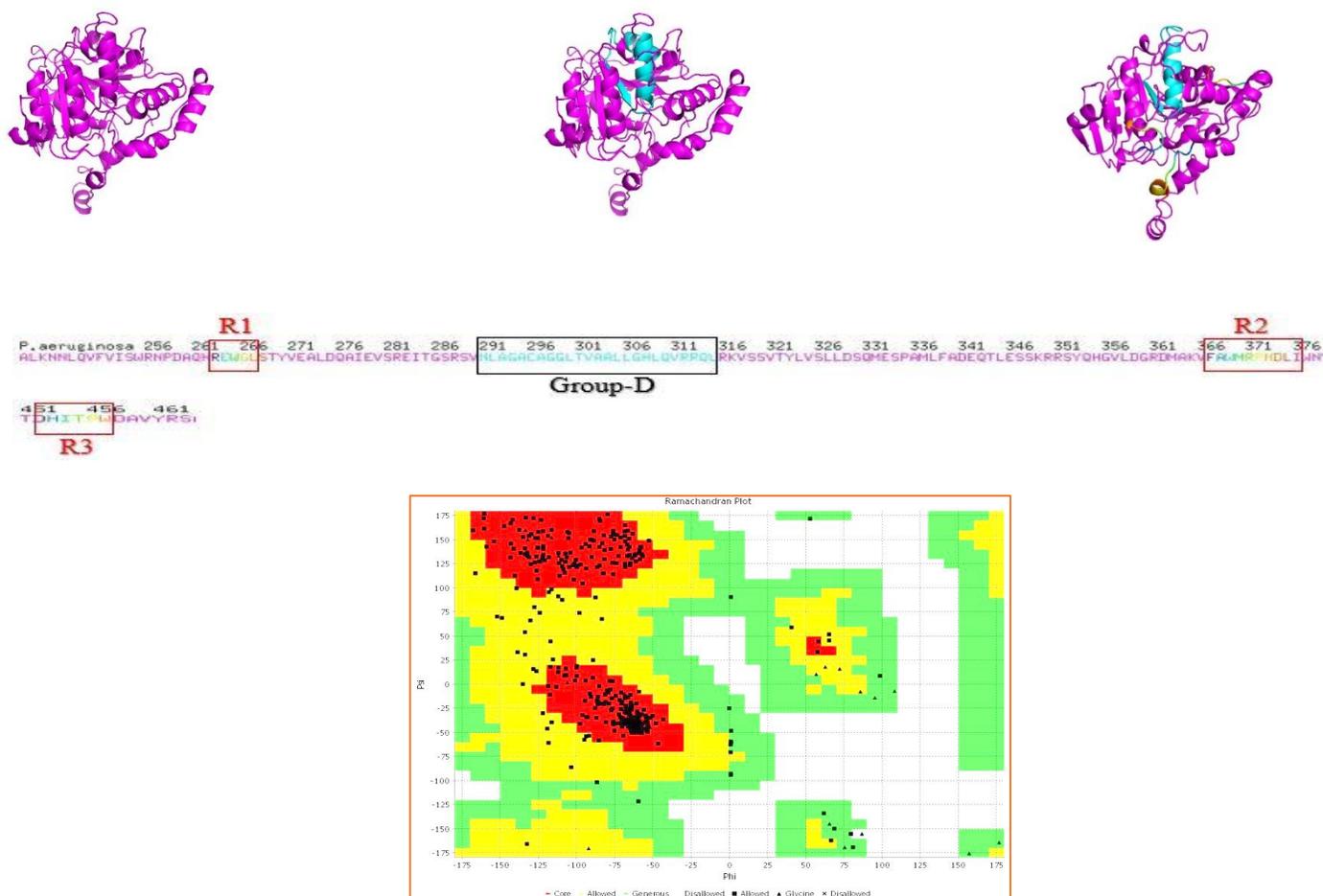


Fig. 10 Evolutionary conserved regions R1, R2, and R3 of PhaC in *Pseudomonas* strain with the group-D conserved site for *P. aeruginosa* POA1. Ramachandran plot for the protein showed that 86.6% of the residues were in favored regions. The graph further indicates that almost 11.2% of amino acids were additionally allowed and 2.1 % in generously allowed regions. No residue of the protein was present in the disallowed region.

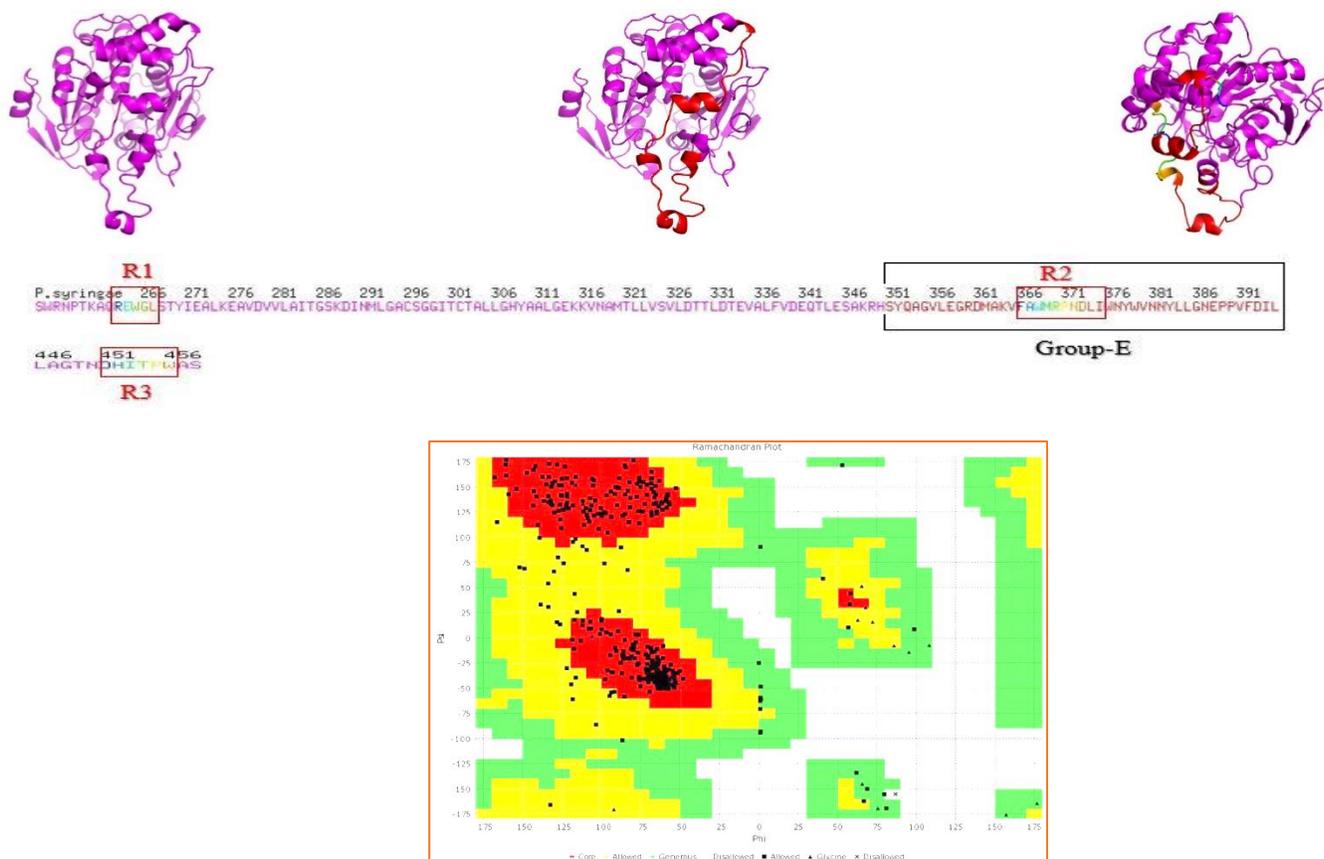


Fig. 11. Evolutionary conserved regions R1, R2, and R3 of PhaC in *Pseudomonas* strain with the group-E conserved site for *P. syringae*. Ramachandran plot for the protein showed that 87.7% of the residues were in favored regions. The graph further indicated that almost 10.2% of amino acids were in additionally allowed and 2.1 % in generously allowed regions. No residue of the protein was present in the disallowed region.

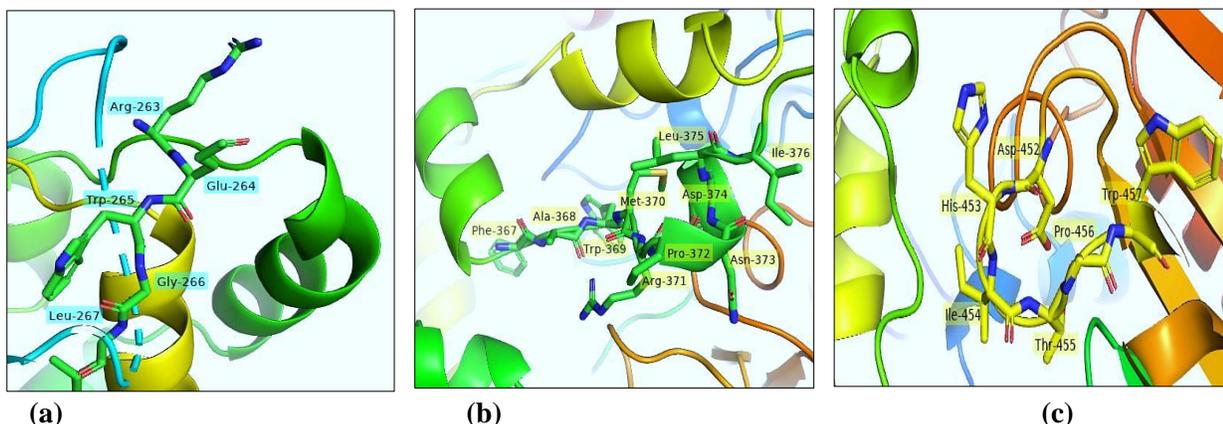


Fig. 12 Evolutionary conserved sites in the PhaC for all the selected strains of *Pseudomonas* (Fig. 1). The highly conserved residues with their locations in the protein chain are shown in (a) Region-1, (b) Region-2, and (c) Region-3. The conserved sites were determined using MEGA7 after aligning the PhaC protein.

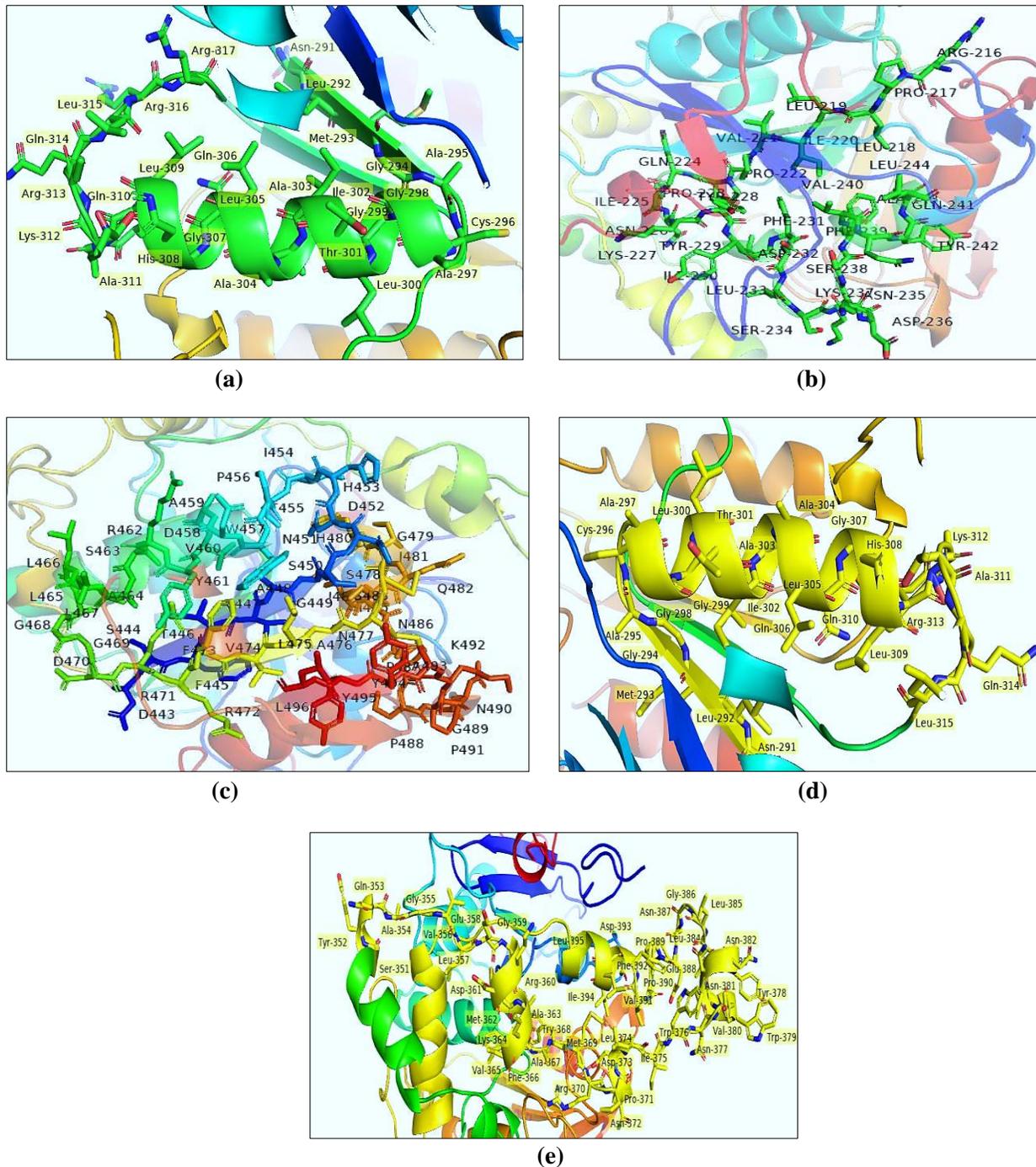


Fig. 13 Evolutionary conserved motifs of PhaC protein among the phylogenetic groups of *Pseudomonas*. The conserved site and its residues can be seen for group-A to group-E as Fig. a-e. These evolutionarily conserved regions for each taxon were determined after aligning the protein sequences using ClustalW, and BLASTp in MEGA7. PyMOL was used to analyze and visualize the amino acid residues corresponding to the sequences.

IV. CONCLUSION

Computational and *in silico* analysis of the forty different strains of *Pseudomonas* for the polyhydroxyalkanoates synthase protein (PhaC) proposed that these strains are evolutionarily related in terms of the conserved regions in the protein. These regions (R1, R2, and R3) have variable numbers of amino acids that might have a crucial role in the structure and function of PhaC. The protein sequence and phylogenetic analysis through various bioinformatics tools confirmed the presence of a unique and identical sequence of amino acids in all the members of phylogenetic groups. It was concluded from the results that *Pseudomonas* strains were experiencing a two-level evolution; one is on the genus level that conserves the same sites in the organisms as a whole while the second is on the specie level where *Pseudomonas* species within a group conserve the same sequence. 3D structural analysis and physiochemical characterization of PhaC protein confirmed a close behavior of the protein that suggested a common evolutionary ancestor. The present study has taken a unique approach towards the analysis of PhaC in terms of not only the structure and conserved sites but also its evolutionary behavior among the *Pseudomonas* species that have never been done before. We believe that further studies on the predicted sites and their possible functions can help molecular scientists design efficient PhaC protein for the hyper-production of PHA leading to new horizons in biopolymer research.

REFERENCES

- [1] F. Muneer, I. Rasul, M. Qasim, A. Sajid, H. Nadeem, Optimization, production and characterization of polyhydroxyalkanoate (PHA) from indigenously isolated novel bacteria, *J. Polym. Environ.* (2022). <https://doi.org/10.1007/s10924-022-02444-y>.
- [2] F. Muneer, I. Rasul, F. Azeem, M.H. Siddique, M. Zubair, H. Nadeem, Microbial polyhydroxyalkanoates (PHAs): Efficient replacement of synthetic polymers, *J. Polym. Environ.* 28 (2020) 2301–2323. <https://doi.org/10.1007/s10924-020-01772-1>.
- [3] P. Sedlacek, E. Slaninova, M. Koller, J. Nebesarova, I. Marova, V. Krzyzanek, S. Obruca, PHA granules help bacterial cells to preserve cell integrity when exposed to sudden osmotic imbalances, *N. Biotechnol.* 49 (2019) 129–136. <https://doi.org/10.1016/j.nbt.2018.10.005>.
- [4] L. Silvestrini, B. Drog, Identification of Four Polyhydroxyalkanoate Structural Genes in *Synechocystis cf. salina* PCC6909: In silico Evidences, *J. Proteomics Bioinform.* 09 (2016) 028–037. <https://doi.org/10.4172/jpb.1000386>.
- [5] P.B.S. Albuquerque, C.B. Malafaia, Perspectives on the production, structural characteristics and potential applications of bioplastics derived from polyhydroxyalkanoates, *Int. J. Biol. Macromol.* 107 (2018) 615–625. <https://doi.org/10.1016/j.ijbiomac.2017.09.026>.
- [6] T. Yamaguchi, J. Narsico, T. Kobayashi, A. Inoue, T. Ojima, Production of poly(3-hydroxybutyrate) by a novel alginolytic bacterium *Hydrogenophaga* sp. strain UMI-18 using alginate as a sole carbon source, *J. Biosci. Bioeng.* 128 (2019) 203–208. <https://doi.org/10.1016/j.jbiosc.2019.02.008>.
- [7] A.H.M.A.S.M.C. Fauzi, L.W. Yoon, T. Nittami, H.K. Yeoh, Enrichment of PHA-accumulators for sustainable PHA production from crude glycerol, *Process Saf. Environ. Prot.* 122 (2019) 200–208. <https://doi.org/10.1016/j.psep.2018.12.002>.
- [8] Muhammadi, Shabina, M. Afzal, S. Hameed, Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: Production, biocompatibility, biodegradation, physical properties and applications, *Green Chem. Lett. Rev.* 8 (2015) 56–77. <https://doi.org/10.1080/17518253.2015.1109715>.
- [9] R. Porta, The Plastics Sunset and the Bio-Plastics Sunrise, *Coatings.* 9 (2019) 526. <https://doi.org/10.3390/coatings9080526>.
- [10] C.G. Avio, S. Gorbi, F. Regoli, Plastics and microplastics in the oceans: From emerging pollutants to emerged threat, *Mar. Environ. Res.* 128 (2017) 2–11. <https://doi.org/10.1016/j.marenvres.2016.05.012>.
- [11] R. Geyer, J.R. Jambeck, K.L. Law, Production, use, and fate of all plastics ever made, *Sci. Adv.* 3 (2017) 25–29. <https://doi.org/10.1126/sciadv.1700782>.
- [12] H. Nadeem, K.B. Alia, F. Muneer, I. Rasul, M.H. Siddique, F. Azeem, M. Zubair, Isolation and identification of low-density polyethylene degrading novel bacterial strains, *Arch. Microbiol.* (2021). <https://doi.org/10.1007/s00203-021-02521-1>.
- [13] J.N. Hahladakis, C.A. Velis, R. Weber, E. Iacovidou, P. Purnell, An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal and recycling, *J. Hazard. Mater.* 344 (2018) 179–199. <https://doi.org/10.1016/j.jhazmat.2017.10.014>.
- [14] F. Muneer, S. Hussain, Sidra-tul-Muntaha, M. Riaz, H. Nadeem, Plastics versus bioplastics, *Materials Research Forum LLC*, 2021. <https://doi.org/10.21741/9781644901335-9>.
- [15] A.K. Singh, L. Sharma, N. Mallick, J. Mala, Progress and challenges in producing polyhydroxyalkanoate biopolymers from cyanobacteria, *J. Appl. Phycol.* 29 (2017) 1213–1232. <https://doi.org/10.1007/s10811-016-1006-1>.
- [16] E. Kabir, R. Kaur, J. Lee, K.H. Kim, E.E. Kwon, Prospects of biopolymer technology as an alternative option for non-degradable plastics and sustainable management of plastic wastes, *J. Clean. Prod.* 258 (2020) 120536. <https://doi.org/10.1016/j.jclepro.2020.120536>.

- [17] F. Muneer, M.H. Azam, M. Zubair, T. Farooq, M. Ibrahim, I. Rasul, M. Afzal, A. Ahmad, H. Nadeem, Remediation of water pollution by plastics, in: Inamuddin (Ed.), *Water Pollut. Remediat. Org. Pollut.*, Springer Nature Switzerland AG 2021, 2021: pp. 89–117. https://doi.org/10.1007/978-3-030-52395-4_3.
- [18] P. Carozzi, A. Giovannelli, M.L. Traversi, E. Touloupakis, T. Di Lorenzo, Poly-3-hydroxybutyrate and H₂ production by *Rhodospseudomonas* sp. S16-VOGS3 grown in a new generation photobioreactor under single or combined nutrient deficiency, *Int. J. Biol. Macromol.* 135 (2019) 821–828. <https://doi.org/10.1016/j.ijbiomac.2019.05.220>.
- [19] M.F. Chek, S.Y. Kim, T. Mori, H.T. Tan, K. Sudesh, T. Hakoshima, Asymmetric open-closed dimer mechanism of polyhydroxyalkanoate synthase PhaC, *IScience.* 23 (2020) 101084. <https://doi.org/10.1016/j.isci.2020.101084>.
- [20] Y. Nambu, M. Ishii-Hyakutake, K. Harada, S. Mizuno, T. Tsuge, Expanded amino acid sequence of the PhaC box in the active center of polyhydroxyalkanoate synthases, *FEBS Lett.* 594 (2020) 710–716. <https://doi.org/10.1002/1873-3468.13651>.
- [21] H.A. Wahab, N.B.A. Khairudin, M.R. Samian, N. Najimudin, Sequence analysis and structure prediction of type II *Pseudomonas* sp. USM 4-55 PHA synthase and an insight into its catalytic mechanism, *BMC Struct. Biol.* 6 (2006) 1–14. <https://doi.org/10.1186/1472-6807-6-23>.
- [22] K. Bin Li, ClustalW-MPI: ClustalW analysis using distributed and parallel computing, *Bioinformatics.* 19 (2003) 1585–1586. <https://doi.org/10.1093/bioinformatics/btg192>.
- [23] T. Ashok Kumar, CFSSP: Chou and Fasman secondary structure prediction server, *Wide Spectr.* 1 (2013) 15–19. <https://doi.org/10.5281/zenodo.50733>.
- [24] V.K. Garg, H. Avashthi, A. Tiwari, P.A. Jain, P.W.R. Ramkete, A.M. Kayastha, V.K. Singh, MFPPi – Multi FASTA ProtParam Interface, *Bioinformation.* 12 (2016) 74–77. <https://doi.org/10.6026/97320630012074>.
- [25] S. Yuan, H.C.S. Chan, Z. Hu, Using PyMOL as a platform for computational drug design, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 7 (2017). <https://doi.org/10.1002/wcms.1298>.
- [26] O. Carugo, K. Djinović-Carugo, A proteomic Ramachandran plot, *Amino Acids.* 44 (2013) 781–790. <https://doi.org/10.1007/s00726-012-1402-z>.
- [27] S.A. Hollingsworth, P.A. Karplus, A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins, *Biomol. Concepts.* 1 (2010) 271–283. <https://doi.org/10.1515/bmc.2010.022>.

AUTHORS

First Author – Ubaid Mehmood, Mater of Philosophy, College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, 2999 Renmin North Road, Shanghai, 201620, P.R. China, ubaidmehmood0@gmail.com

Second Author – Faizan Muneer, Master of Philosophy, Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Faisalabad, Punjab-Pakistan. drfm13434@gmail.com

Third Author – Xiao Junhua, Ph.D. College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, 2999 Renmin North Road, Shanghai, 201620, P.R. China. xiaojunhua@dhu.edu.cn

Correspondence Author – Habibullah Nadeem, habibullah@gcuf.edu.pk

