

COMPARATIVE AMINO ACIDS STUDIES ON PHAC SYNTHASES AND PROTEASES AS WELL AS ESTABLISHING A NEW TREND IN EXPERIMENTAL DESIGN

AMRO A. AMARA^{1,2,3}, SALEH M. MATAR¹ AND EHAB SEROUR^{1,3}

¹*Genetic Engineering and Biotechnology Research Institute,
Mubarak City for Scientific Research and Technology Applications,
Alexandria, Egypt.*

²*Microbiology Division, Pharmaceutics Department, College of Pharmacy,
King Saud University, Riyadh, Kingdom of Saudi Arabia.*

³*King Abdulaziz City for Science and Technology, Riyadh,
Kingdom of Saudi Arabia.*

amroamara@web.de

ABSTRACT: A question addressed in this study is: why similar enzymes are classified into different subclasses? As an example, PhaC synthases are classified according to four different classes (I, II, III and IV). To answer this question we proposed that besides the catalytic residues, the overall amino acids (AAs) present are responsible for the differences observed. The AAs' composition affects the structure/function/substrate specificity (SFS) of these enzymes. The differences between the classes in various PhaC synthases and proteases were analysed to support our argument. Homology and phylogenetic tree of some selected PhaC synthases of different strains (representing the four classes) were demonstrated. The properties of a specific class of enzyme could not be changed into those of another by changing the catalytic residues. Moreover, these differences could not be detected from the proteins' 3D structures, despite clear differences at the AAs level. Another question was also addressed: could we benefit from the various existing protein databases in the field of biotechnology? To answer this, we introduced a model for an Experimental Design based on the information in the protein database (for strains available in our lab) regarding their ability to degrade castor oil. Two enzymes in the phenol degradation pathway, phenol 2-monooxygenase and catechol 1,2-dioxygenase, and a lipase enzyme were analysed. These enzymes were screened and analysed according to the BLAST-protein database and BRENDA. The comprehensive enzyme information system compared six strains against each other, including: *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Bacillus pumilus*, *Bacillus thuringiensis*, *Bacillus licheniformis*, and *Geobacillus stearothermophilus*. Only *P. aeruginosa* proved to have the three required enzymes and was suitable for the production of lipases from castor oil (crude castor oil is usually contaminated with phenol) as indicated by the databases. In addition, *in vivo* castor oil degradation and *in vitro* lipase enzyme activity were analysed. The apparent lipase activity was 1070 Units/ml. Therefore, this new strategy is recommended to better understand the SFS as well as for using protein database in an Experimental Design.

ABSTRAK: Kajian ini berkisar mengenai soal: mengapa enzim yang sama diklasifikasikan kepada subkumpulan yang berbeza? Contohnya sintesis PhaC dikelaskan kepada empat kumpulan berbeza (I, II, III dan IV). Bagi menjawab soal ini selain hipotesis katalitik residue diperkenalkan, faktor yang bertanggungjawab adalah asid amino (AAs) secara keseluruhannya. Komposisi AAs memberi kesan kepada struktur/fungsi/spesifikasi substrat (SFS) enzim-enzim ini. Perbezaan di antara kumpulan dalam pelbagai sintesis PhaC dan proteases telah dianalisis bagi menyokong hujah ini. Homologi dan asas phylogenetic bagi sintesis PhaC tertentu yang berbeza *strains*

(mewakili empat kumpulan) telah ditunjukkan. Sifat-sifat tidak boleh ditukarkan dengan menukarkan katalitik residue dari satu kepada yang lain. Tambahan pula, kelainannya tidak boleh dikesan dari struktur 3D protein, walaupun perbezaan yang nyata pada peringkat AAs. Soalan lain yang berkisar adalah: Adakah kita mendapat faedah dari pelbagai pengkalan data dalam bidang bioteknologi? Bagi menjawab soalan ini, model Rekabentuk Eksperimen telah diperkenalkan yang berasaskan maklumat pengkalan data protein bagi *strains* di makmal kami yang boleh degradasi minyak kastor. Dua enzim dalam degradasi fenol telah dianalisis, fenol 2-monooxygenas dan catechol 1,2-dioxygenas serta enzim lipas. Enzim-enzim ini telah disaringkan dan dianalisis dengan merujuk kepada pengkalan data protein-BLAST dan BRENDA. Sistem maklumat enzim secara komprehensif terhadap enam *strains* termasuk *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Bacillus pumilus*, *Bacillus thuringiensis*, *Bacillus licheniformis* dan *Geobacillus stearothermophilus*. Hanya *P. Aeruginosa* terbukti mempunyai tiga enzim dan sesuai bagi penghasilan *lipase* dari minyak kastor (minyak kastor mentah selalunya tercemar dengan fenol) seperti yang ditunjukkan dari pengkalan data. Degradasi minyak kastor *in vivo* dan aktiviti enzim *lipase in vitro* telah dilaksanakan. Aktiviti *lipase* jelas adalah 1070 Units/ml. Kami mencadangkan menggunakan strategi ini bagi memahami SFS serta pengkalan data protein dalam Rekabentuk Eksperimen.

KEYWORDS: *amino acids; model; PhaC synthase; protease; lipase; experimental design*

1. INTRODUCTION

Minor differences in the enzymes' AAs constituents could lead to classifying certain group of enzymes to different classes [1, 2]. PhaC synthases are included in one group of enzymes responsible for the production of various types of the bioplastics-PHAs [3]. PhaC synthases have been classified to four classes: I, II, III and IV [3]. The produced polymers from those classes can be classified into three different types: short chain length (SCL), medium chain length (MCL) and long chain length (LCL). In between these polymers there are a series of mix/co-polymer of all types [3].

Proteases which are our second enzyme have been classified to larger number of classes [2]. Rehm et al. [4], Amara et al. [4, 5] and Taguchi et al. [7, 8] showing and proving that mutants in places other than the PhaC synthases (catalytic residues) active sites could influence their activities [4-8]. They have used quite different protocols for conducting such mutations [4-8]. This clearly shows that the amino acids composition of certain enzyme are involved in the protein activity beside the catalytic ones. In the EMBO conferences for comparative genomics of Eukaryotic microorganisms (2009) Amara et al. [1] and Serour et al. [2] show a new trend in evaluating the different classes of one type of enzyme based on the AAs constituents [1, 2]. The examples were focused on PhaC synthases found in Class I, II, III and IV from microbes and algae as well as proteases from different microbes and plants [1, 2]. Mater et al. [9] and for the first time represent a tool for implementing different nucleotide and protein databases for improving the design of the lipase production experiment [9]. Recently, Amara and Salem [10] showed an efficient protocol for lipase production from pharmaceutical grade castor oil using *Pseudomonas aeruginosa* [10]. Moreover, Amara [11], show the possibility for production of lipases, PHA and mesophilic proteases simultaneously using a well thought experimental design [11]. In this study a deeper investigation for the role of AAs profiles in the determination and influencing the enzymes activities and structures have been conducted. A simpler tool for establishing experimental design based on DNA/protein database has been established

which can improve both of the *in vivo* enzyme production and increasing the *in vitro* activities.

2. MATERIALS AND METHODS

2.1 Data Collection

For investigating the AAs differences between PhaC synthases in different classes - I, II, III and IV - one hundred sequences from each of the first three classes have been collected from BLAST-protein database and one sequence from class IV. The sequences of each class have been saved in text file in form of FASTA format. The same process has been followed in case of proteases enzymes where seven classes have been used and represent alkaline proteases (12 sequences); *Aspergillus* proteases (20 sequences); *Candida* proteases (18 sequences); mesophilic proteases (7 sequences); metallo-proteases (20 sequences); serine proteases (24 sequences) and plant proteases (7 sequences). The overall sequences used have been described by their numbers and names/or reduced names as given below.

2.1.1 PhaC synthases (Class I)

I-73541041-Ralstonia eutropha JMP134; I-113867451-Ralstonia-eutropha H16588; I-58045497-Cupriavidus necator; I-194289467-Cupriavidus taiwanensis; I-94310298-Ralstonia metallidurans CH34; I-83745885-Ralstonia solanacearum UW551; I-207724165-Ralstonia solanacearum MolK2; I-17546350-Ralstonia solanacearum GMI1000; I-187928996-Ralstonia pickettii 12J; I-241663184-Ralstonia pickettii 12D; I-6683779-Burkholderia sp. DSMZ 9242; I-171058770-Leptothrix cholodnii SP-6; I-254252278-Burkholderia dolosa AU0158; I-121605699-Polaromonas naphthalenivorans CJ2; I-238027092-Burkholderia glumae BGR1; I-239815023-Variovorax paradoxus S110; I-221198057-Burkholderia multivorans CGD2M; I-161524656-Burkholderia multivorans ATCC 17616; I-21218117-Pseudomonas putida; I-115351780-Burkholderia ambifaria AMMD; I-221212792-Burkholderia multivorans CGD1; I-186911857-Burkholderia sp. Rc07; I-254248091-Burkholderia cenocepacia PC184; I-206560226-Burkholderia cenocepacia J2315; I-172060751-Burkholderia ambifaria MC40-6; I-76809230-Burkholderia pseudomallei 1710b; I-53719168-Burkholderia pseudomallei K96243; I-107029029-Burkholderia cenocepacia AU 1054; I-167902422-Burkholderia pseudomallei NCTC 13177; I-167823894-Burkholderia pseudomallei 9; I-167719296-Burkholderia pseudomallei DM98; I-237812525-Burkholderia pseudomallei MSHR346; I-167836425-Burkholderia thailandensis MSMB43; I-170733151-Burkholderia cenocepacia MC0-3; I-126451661-Burkholderia pseudomallei 1106a; I-134277727-Burkholderia pseudomallei 305; I-126441293-Burkholderia pseudomallei 668; I-171322801-Burkholderia ambifaria MEX-5; I-167893973-Burkholderia pseudomallei 7894; I-170703956-Burkholderia ambifaria IOP40-10; I-167589915-Burkholderia ubonensis Bu; I-53723502-Burkholderia mallei ATCC 23344; I-67641326-Burkholderia mallei GB8 horse 4; I-74486339-Pseudomonas sp. 10c-1-3; I-241766811-Acidovorax delafieldii 2AN; I-1688262-Alcaligenes sp.; I-167581728-Burkholderia thailandensis TXDOH; I-134295822-Burkholderia vietnamiensis G4; I-167562575-Burkholderia oklahomensis EO147; I-186475734-Burkholderia phymatum STM815; I-83720440-Burkholderia thailandensis E264; I-78066561-Burkholderia sp. 383; I-167619844-Burkholderia thailandensis Bt4; I-91788012-Polaromonas sp. JS666; I-170692350-Burkholderia graminis C4D1M; I-187923833-Burkholderia phytofirmans PsJN; I-256759319-Thiomonas intermedia K12; I-91783473-Burkholderia xenovorans LB400; I-209520555-Burkholderia sp. H160; I-124267059-Methylobium petroleiphilum PM1; I-89901339-Rhodospirillum rubrum T118; I-4097186-Alcaligenes latus; I-260219581-Curvibacter putative symbiont of Hydra magnipapillata; I-120611148-Acidovorax avenae subsp. citrulli AAC00-1; I-3982781-Alcaligenes latus; I-121610739-Verminophobacter eiseniae EF01-2; I-126032017-uncultured bacterium; I-222110875-Diaphorobacter sp. TPSY; I-121594430-Acidovorax sp. JS42; I-145589106-Polynucleobacter necessarius subsp. asymbioticus QLW-P1DMWA-1; I-134095122-Herminiimonas arsenicoxydans; I-152980974-Janthinobacterium sp. Marseille; I-171463618-Polynucleobacter necessarius subsp. necessarius STIR1; I-1513226-Zoogloea ramigera; I-163856159-Bordetella petrii DSM 12804; I-33592281-Bordetella pertussis Tohama I; I-33597723-Bordetella parapertussis 12822; I-33602571-Bordetella bronchiseptica RB50; I-149926148-Limnobacter sp. MED105; I-71907240-Dechloromonas aromatica RCB; I-237653784-Thauera sp. MZ1T; I-68164161-Burkholderia sp. BCC6; I-187477719-Bordetella avium 197N; I-27378843-Bradyrhizobium japonicum USDA 110; I-94310611-Ralstonia metallidurans CH34; I-226944453-Azotobacter vinelandii DJ; I-56478108-Aromatoleum aromaticum EbN1; I-119897312-Azoarcus sp. BH72; I-14456069-Azotobacter sp. FA8; I-186471503-Burkholderia phymatum STM815; I-157168636-Pseudomonas sp. PC17; I-4062965-Pseudomonas sp. 61-3; I-126032011-uncultured bacterium; I-224825073-Lutielia nitroferrum 2002; I-54306501-Pseudomonas oleovorans; I-117165349-Pseudomonas sp. 14-3; I-146281078-Pseudomonas stutzeri A1501; I-223711877-Pseudomonas sp. USM4-55; I-89093191-Oceanospirillum sp. MED92; I-258654984-Nakamurella multipartita DSM 44233

2.1.2 *PhaC synthases (Class II)*

II-170719647-Pseudomonas putida W619; II-30909309-Pseudomonas putida; II-219523978-Pseudomonas putida; II-167036042-Pseudomonas putida GB-1; II-129993-; II-148550079-Pseudomonas putida F1; II-26991680-Pseudomonas putida KT2440; II-104783984-Pseudomonas entomophila L48; II-52355297-Pseudomonas putida; II-55140655-Aeromonas hydrophila; II-21689573-Pseudomonas putida; II-15028442-Burkholderia caryophylli; II-224593170-Pseudomonas chlororaphis; II-229587957-Pseudomonas fluorescens SBW25; II-70733938-Pseudomonas fluorescens Pf-5; II-55925768-Comamonas testosterone; II-55925776-Pseudomonas sp. KBOS 04; II-17402511-Pseudomonas chlororaphis subsp. Aureofaciens; II-71736177-Pseudomonas syringae pv. phaseolicola 1448A; II-4062968-Pseudomonas sp. 61-3; II-213969431-Pseudomonas syringae pv. tomato T1; II-62637851-Pseudomonas corrugate; II-118200329-Pseudomonas corrugate; II-223470533-Pseudomonas sp. USM7-7; II-55925780-Pseudomonas sp. KBOS 17; II-218137968-Pseudomonas fluorescens; II-62637857-Pseudomonas mediterranea; II-77456622-Pseudomonas fluorescens Pf0-1; II-55925772-Pseudomonas sp. KBOS 03; II-256709422-Pseudomonas sp. gl06; II-222840540-Pseudomonas sp. MBEL 6-19; II-54610882-Pseudomonas sp. 3Y2; II-161344526-Pseudomonas sp. USM4-55; II-83752359-type II Pseudomonas mendocina; II-28916412-Pseudomonas stutzeri; II-146305567-Pseudomonas mendocina ymp; II-13346169-Pseudomonas pseudoalcaligenes; II-19589603-Pseudomonas nitroreducens; II-15421135-Pseudomonas pseudoalcaligenes; II-34452167-Pseudomonas sp. HJ-2; II-10835919-Pseudomonas resinovorans; II-20086523-Pseudomonas mendocina; 559 II-152988584-Pseudomonas aeruginosa PA7; II-107104156-Pseudomonas aeruginosa PACS2; II-15600249-Pseudomonas aeruginosa PA01; II-116053204-Pseudomonas aeruginosa UCBBP-PA14; II-254238237-Pseudomonas aeruginosa C3719; II-54610876-Pseudomonas sp. 3Y2; II-20069348-synthetic construct; II-20069350-synthetic construct; II-34452169-PhaC2 Pseudomonas sp. HJ-2; II-218894157-PhaC2 Pseudomonas aeruginosa LESB58; II-116053206-PhaC2 Pseudomonas aeruginosa UCBBP-PA14; II-107104158-Pseudomonas aeruginosa PACS2; II-10835920-PhaC2 Pseudomonas resinovorans; II-15600251-PhaC2 Pseudomonas aeruginosa PA01; II-254238235-PhaC2 Pseudomonas aeruginosa C3719; II-152984502-PhaC2 Pseudomonas aeruginosa PA7; II-146305565-synthase, II Pseudomonas mendocina ymp; II-17402513-PhaC2 Pseudomonas chlororaphis subsp. Aureofaciens; II-28916414-PhaC2 Pseudomonas stutzeri; II-55925774-PhaC2 Pseudomonas sp. KBOS 03; II-54610884-PhaC2 Pseudomonas sp. 3Y2; II-13346173-PhaC2 Pseudomonas nitroreducens; II-77456620-Pseudomonas fluorescens Pf0-1; II-15421137-PhaC2 Pseudomonas pseudoalcaligenes; II-78062240-Burkholderia sp. 383; II-70733936-Pseudomonas fluorescens Pf-5; II-55925782-PhaC2 Pseudomonas sp. KBOS 17; II-104783986-polymerase 2 Pseudomonas entomophila L48; II-256709424-PhaC2 Pseudomonas sp. gl06; II-4062970-PhaC2 Pseudomonas sp. 61-3; II-62637859-PhaC2 Pseudomonas mediterranea; II-255957404-PhaC2 Pseudomonas extremaustralis; II-218137970-Pseudomonas fluorescens; II-62637853-PhaC2 Pseudomonas corrugate; II-161344527-synthase C2 Pseudomonas sp. USM4-55; II-229587955-polymerase 2 Pseudomonas fluorescens SBW25; II-182638936-PhaC2 Pseudomonas mendocina; II-37925926-PhaC Pseudomonas fluorescens; II-51234091-PhaC2 Pseudomonas corrugate; II-26991682-polymerase 2 Pseudomonas putida KT2440; II-148550081-Pseudomonas putida F1; II-15028444-PhaC2 Burkholderia caryophylli; II-55925778-PhaC2 Pseudomonas sp. KBOS 04; II-20086525-PhaC2 Pseudomonas mendocina; II-223470535-PhaC2 Pseudomonas sp. USM7-7; II-130005-polymerase 2; II-21689575-PhaC2 Pseudomonas putida; II-257482918-Pseudomonas syringae pv. tabaci ATCC 11528; II-71736818-Pseudomonas syringae pv. phaseolicola 1448A; II-167036044-Pseudomonas putida GB-1; II-6049268-PhaC2 Pseudomonas putida; II-28872257-Pseudomonas syringae pv. tomato str. DC3000; II-52355299-polymerase 2 Pseudomonas putida; II-4104927-Pseudomonas putida; II-55925770-PhaC2 Comamonas testosterone; II-170719645-Pseudomonas putida W619; II-213969429-Pseudomonas syringae pv. tomato T1; II-30909311-PhaC2 Pseudomonas putida;

2.1.3 *PhaC synthases (Class III)*

III-52144019-Bacillus cereus E33L; III-30261419-Bacillus anthracis str. Ames; III-94538301-Bacillus cereus; III-229144020-Bacillus cereus BDRD-ST24; III-228938531-Bacillus thuringiensis serovar berliner ATCC 10792; III-229043156-Bacillus cereus AH676; III-47566204-Bacillus cereus G9241; III-218230992-Bacillus cereus B4264; III-75762375-Bacillus thuringiensis serovar israelensis ATCC 35646; III-30019469-Bacillus cereus ATCC 14579; III-229189498-Bacillus cereus ATCC 10876; III-125962439-Bacillus cereus; III-229102015-Bacillus cereus Rock3-28; III-228907048-Bacillus thuringiensis IBL 200; III-229160368-Bacillus cereus R309803; III-229095902-Bacillus cereus Rock3-29; III-229029086-Bacillus cereus AH1271; III-163939223-Bacillus weihenstephanensis KBAB4; III-229172055-Bacillus cereus MM3; III-229016661-Bacillus cereus AH1273; III-229010712-Bacillus mycoides DSM 2048; III-229084404-Bacillus cereus Rock3-44; III-228990419-Bacillus pseudomycoideus DSM 12442; III-152974848-Bacillus cereus subsp. cytotoxis NVH 391-98; III-4160475-Bacillus megaterium; III-214028400-Bacillus sp. C18; III-214028405-Bacillus sp. E13; III-108805188-Rubrobacter xylanophilus DSM 9941; III-156740130-Roseiflexus castenholzii DSM 13941; III-114566626-Syntrophomonas wolfei subsp. wolfei str. Goettingen; III-148658629-Roseiflexus sp. RS-1; III-214028407-Bacillus sp. C19; III-114566767-wolfei subsp. wolfei str. Goettingen; III-219847569-Chloroflexus aggregans DSM 9485; III-163848796-Chloroflexus aurantiacus J-10-fl; III-227882098-Halogeometricum borinquense DSM 11551; III-170177749-Haloferax mediterranei; III-27377996-Bradyrhizobium japonicum USDA 110; III-254446588-subfamily Verrucomicrobiae bacterium DG1235; III-227408209-Geodermatophilus obscurus DSM 43160; III-223938536-subunit bacterium Ellin514; III-108757578-subunit Myxococcus xanthus DK 1622; III-119474707-marine gamma proteobacterium HTCC2143; III-55379489-Haloarcula marismortui ATCC 43049; III-257053230-Halorhabdus utahensis DSM 12940; III-157779733-Haloarcula hispanica; III-146275777-Novosphingobium aromaticivorans DSM 12444; III-220934684-subunit Thioalkalivibrio sp. HL-

EbGR7; III-161528499-Nitrosopumilus maritimus SCM1; III-89100299-Bacillus sp. NRRL B-14911; III-257387054-Halomicrobium mukohataei DSM 12286; III-209522880-Arthrospira maxima CS-328; III-110668241-Haloquadratum walsbyi DSM 16790; III-18642978-Synechococcus sp. MA192; III-118576513-Cenarchaeum symbiosum A; III-260571529-Ferroglobus placidus DSM 10642; III-218437501-Cyanothece sp. PCC 7424; III-166367744-Microcystis aeruginosa NIES-843; III-124008767-Microscilla marina ATCC 23134; III-159029277-Microcystis aeruginosa PCC 7806; III-241776928-Allochrochromatium vinosum DSM 180;355 III-1172474-; III-167043660-marine crenarchaeote HF4000_APKG2016; III-89093232-Oceanospirillum sp. MED92; III-16330022-Synechocystis sp. PCC 6803; III-1172475-; III-118132680-Allochrochromatium vinosum; III-18644660-Chlorogloeopsis fritschii; III-220909421-Cyanothece sp. PCC 7425; III-11096252-Ectothiorhodospira shaposhnikovii; III-257065631-Anaerococcus prevotii DSM 20548; III-117924470-Magnetococcus sp. MC-1; III-188576604-Xanthomonas oryzae pv. oryzae PXO99A; III-58581844-Xanthomonas oryzae pv. oryzae KACC10331; III-116254554-Rhizobium leguminosarum bv. viciae 3841; III-78047765-Xanthomonas campestris pv. vesicatoria str. 85-10; III-166712341-Xanthomonas oryzae pv. oryzicola BLS256; III-224371266-Desulfobacterium autotrophicum HRM2; III-21242792-Xanthomonas axonopodis pv. citri str. 306; III-21231598-Xanthomonas campestris pv. campestris str. ATCC 33913; III-254521956-Stenotrophomonas sp. SKA14; III-190574867-Stenotrophomonas maltophilia K279a; III-194366191-Stenotrophomonas maltophilia R551-3; III-114321632-Alkalilimnicola ehrlichii MLHE-1; III-22797878-uncultured crenarchaeote; III-257094743-Candidatus Accumulibacter phosphatis clade IIA str. UW-1; III-170748818-subunit Methylobacterium radiotolerans JCM 2831; III-88811524-Nitrococcus mobilis Nb-231; III-170739114-Methylobacterium sp. 4-46; III-260571530-Ferroglobus placidus DSM 10642; III-120556700-Marinobacter aquaeolei VT8; III-110834279-Alcanivorax borkumensis SK2; III-254429958-Alcanivorax sp. DG881; III-184158035-Acinetobacter baumannii ACICU; III-83644584-Hahella chejuensis KCTC 2396; III-193077315-Acinetobacter baumannii ATCC 17978; III-260555086-Acinetobacter baumannii ATCC 19606; III-169796045-Acinetobacter baumannii AYE; III-169633532-Acinetobacter baumannii SDF; III-260553937-Acinetobacter sp. RUH2624;

2.1.4 *PhaC synthases (Class IV)*

IV-Bacillus megaterium.

2.1.5 *Alkaline proteases*

ALK-13562134-thermophilic alkaline; ALK-257891280-thermophilic alkaline; ALK-56378587-alkaline serine protease; ALK-56418745-alkaline serine protease; ALK-401729-alkaline serine proteinase; ALK-11408; ALK-62906871-serine proteinase; ALK-56910138-Bacillus clausii KSM-K16; ALK-56908597-aminopeptidase T-Bacillus; ALK-56963895-aminopeptidase-Bacillus; ALK-56962360-aminopeptidase T-Bacillus; ALK-142532-subtilisin J-Geobacillus sp.

2.1.6 *Aspergillus proteases*

ASP-470729-Aspergillus niger; ASP-32693448-Aspergillus oryzae; ASP-73921582; ASP-94730401; ASP-217809-alkaline protease-Aspergillus; ASP-2428-precursor alkaline protease; ASP-119399000-Aspergillus clavatus; ASP-119396560-Aspergillus clavatus; ASP-115384910-Aspergillus terreus; ASP-71001650-serine protease Pim1; ASP-70997972-serine protease Alp2; ASP-7099; ASP-2785-Aspergillus fumigates; ASP-285611-Aspergillus oryzae; ASP-114196694-Aspergillus terreus; ASP-244499-alkaline protease Alp; ASP-234833-neutral protease II NP11; ASP-119404410-Aspergillus clavatus NRRL; ASP-119403619-Aspergillus clavatus NRRL; 119403502-putative-Aspergillus clavatus.

2.1.7 *Candida proteases*

CA-1168762;CA-115618; CA-1168764; CA-1705600;CA-68053267;CA-1168768;CA-3334136;CA-1168766;CA-3334137;CA-1168771;CA-1168770;CA-14488057-subtilisin-like-PRB1;CA-578123-aspartyl protease-Candida; CA-456252-aspartyl protease-Candida;CA-456249-aspartyl protease-Candida;CA-220900311-aspartyl protease-Candida;CA-220900309-secreted aspartyl;CA-220900307-secreted aspartyl;CA-220900305-secreted aspartyl;CA-220900303-secreted aspartyl.

2.1.8 *Mesophilic proteases*

M-151567838-Thermobifida Fusca; M-151567837-Thermobifida Fusc; M-146386486-Psychrophilic; M-11345441-Acinetobacter sp; M-135017-Alkaline mesentericopeptidase; M-159794853-Sphericase; M-159794852-Sphericase.

2.1.9 *Metallo-proteases*

ML-120615033-Bartonella;ML-110801468-Clostridium;ML-121602856-Bartonella;ML-110800906-Clostridium;ML-256258533-Brucella;ML-110681969-Clostridium;ML-110675553-Clostridium;ML-254731309-Brucella abortus;ML-254696395-Brucella abortus;ML-254694768-Brucella abortus;ML-254690280-Brucella abortus;ML-254478089-Carboxydibrachium;ML-254446421-Verrucomicrobiae;ML-189025198-Brucella abortus; ML-214035951-Carboxydibrachium;ML-198260729-Verrucomicrobiae;ML-

189020770-Brucella abortus;ML-182625020-Clostridium;ML-182381872-Clostridium;ML-168216254-Clostridium.

2.1.10 Serine proteases

SE-56419065-serine protease; SE-22299979-serine protease; SE-22296164-serine protease; SE-147679133-serine protease; SE-147679078-serine protease; SE-58176722-A Thermophilic;SE-56381846-Geobacillus;SE-56378907-serine protease; SE-56378709-serine protease; SE-56422004-serine protease; SE-56418867-serine protease; SE-212640652-Trypsin-like; SE-212639018-Serine protease; SE-22297602-putative serine protease; SE-212562132-Trypsin-like protein; SE-212560498-Serine protease; 56378587-alkaline serine protease; SE-212640455-subtilisin-like; SE-212637927-ATP-dependent serine; SE-146275230-hypothetical serine; SE-146275175-hypothetical serine; SE-146272717-ATP-dependent serine; SE-212561935-subtilisin-like serine; SE-212560214-Serine protease.

2.1.11 Plant proteases

PL-3859607-cysteine proteases; PL-3859606-cysteine proteases; PL-6137413-Ginger Rhizome Zingiber; PL-6137412-Ginger Rhizome Zingiber; PL-6137410-Ginger Rhizome Zingiber; PL-18141285-cysteine protease; PL-47524507-putative cysteine.

2.2 Generating AAs Profiles

For each of the four different classes of PhaC synthases and each of the seven proteases, AAs profiles were generated. The software OMGA 0.4 was used to analyse the sequences collected for each class individually. OMGA 0.4 enables to calculate the % of each AAs in each protein and gives an average for the used proteins within each class. The average AAs % of the twenty AAs for each class has been collected and summarized in Table 1.

2.3 Generating a Sample Phylogenic Tree and AAs Alignment

Sample phylogenic tree and an alignment for the proteins primer sequences for selected PhaC synthases have been generated. The phylogenic tree and the sequences alignment have been generated for the PhaC synthases different classes using BioEdit v. 7.0.9.0. and MEGA 4.

2.4 Generating PhaC Synthases Protein Models

A Model for each of the four classes of the PhaC synthases has been generated. The software Modeller v 9.8 was used for the models generation. The models have been aligned to each other to show the similarity and the differences between their structures. *Ralstonia eutropha* was used to represent Class I, *Pseudomonas putida* was used to represent Class II, *Bacillus thuringiensis* was used to represent Class III and *Bacillus megaterium* was used to represent ClassIV.

2.5 Database/Biotechnology Experimental Design

2.5.1 Microbial Strains

Pseudomonas aeruginosa, *Bacillus subtilis*, *Bacillus pumilus*, *Bacillus thuringiensis*, *Bacillus licheniformis* and *Geobacillus stearothermophilus* which are available in our lab have been investigated using protein database (BLAST) regarding to their ability to produce lipases from castor oil and their ability to survive in the presence of the phenol.

2.5.2 Investigating Essential Enzymes

Three essential enzymes have been investigated. They are the first two enzymes in the phenol degradation pathway: EC 1.14.13.7 - Phenol-2 monooxygenase and EC. 1.13.11.1 - Catechol 1,2-dioxygenase (BRENDA-The comprehensive enzyme information system) as

well as a lipase enzyme. These three enzymes have been analyzed against BLAST-protein database and against the above described microbial strains.

Table 1: PhaCs and proteases profiles of AAs

Proteins	Amino Acids †																Total				
	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser		Thr	Val	Irp	Tyr
PhaC synthases, Class I (100 sequences)	11.63	0.82	5.40	4.91	4.12	7.16	1.76	3.79	4.03	9.88	2.26	4.18	6.62	4.99	4.94	6.10	5.31	6.07	2.84	3.20	591.3
PhaC synthases, Class II (100 sequences)	8.98	0.82	5.41	4.48	3.64	6.70	2.78	3.67	4.45	11.59	1.98	6.03	5.86	4.31	5.81	6.75	5.00	6.14	2.66	2.95	559.2
PhaC synthases, Class III (100 sequences)	6.07	0.83	7.26	6.02	3.97	6.63	1.92	5.83	5.68	10.96	2.98	4.28	5.31	5.05	5.05	4.97	4.95	7.36	1.72	4.63	364.9
PhaC synthases, Class IV (1 sequences)	6.35	0.28	6.91	5.80	4.14	5.52	1.10	6.35	9.12	9.94	3.31	4.97	5.80	4.14	4.14	5.52	4.70	6.35	1.93	5.52	362
Alkaline proteases (12 sequences)	10.49	0.76	5.94	5.34	3.05	9.14	2.18	5.09	5.22	7.59	2.20	4.96	4.37	2.82	4.26	7.25	6.64	7.72	1.31	3.67	393.1
Aspergillus Proteases (20 sequences)	8.89	1.04	6.14	6.56	3.19	7.53	1.99	5.16	5.79	8.07	1.53	3.86	5.58	3.64	5.25	8.53	6.07	7.36	0.92	2.91	582.3
Candida Proteases (18 sequences)	6.60	1.04	7.33	3.14	4.34	8.31	1.06	6.76	5.79	7.99	0.80	6.53	3.90	4.23	2.24	10.78	6.83	7.20	0.70	4.43	419.4
Mesophilic Proteases (7 sequences)	11.08	0.99	4.65	2.72	2.09	12.96	1.52	6.17	2.98	5.43	0.94	5.80	3.45	4.28	3.29	10.97	8.31	6.95	1.36	4.08	273.4
Metallo Proteases (20 sequences)	10.21	0.96	5.88	8.49	3.92	8.75	2.48	5.64	6.16	8.19	2.21	4.41	3.63	1.93	5.05	5.91	4.63	6.45	2.16	2.95	394.4
Plant Proteases (7 sequences)	6.77	3.54	5.86	5.71	4.09	9.39	1.72	5.76	5.81	6.21	1.36	6.72	4.34	3.43	3.28	8.18	5.05	7.12	2.17	3.48	282.9
Serine Proteases (24 sequences)	10.07	0.44	5.63	5.93	3.42	9.00	1.49	5.64	5.35	8.47	2.22	3.84	5.61	3.32	4.58	6.05	5.03	9.25	1.13	3.53	519.8

These classes are: Class I, II, III and IV. Comparative studies between the PhaC synthases within different classes has been performed in this study regarding to their AAs constituents and the alignment of their different primary structures aiming to understand why these enzymes show significant differences in substrate specificity and the type of polymer(s) produced.

Proteases which are one of the most important technical enzymes with significant market capacity have been investigated too. PhaC synthases have been selected as a simple example while seven “classes” of proteases have been selected as a more diverse and complex example. An early study by Amara et al. [6] show improvement in the PhaC synthases activities due to mutations generated in AAs other than the catalytic residues [6]. These mutants lead to enhanced enzyme activities. This agrees with our argument in this study. AAs within the overall protein backbone are important in SFS. These mutations did not only improve the enzymes activities but also change their substrate specificity which enables incorporation of different monomeric constituents and give new polymers. In fact during Amara et al. [6] study only the enzymes which show enhanced activities have been highlighted, however many of those which have less activities have been neglected [6]. This is an important illustration that the protein active sites are not the single source that affect and influence SFS properties. Even the study conducted by Amara et al. [6] agrees with that but here a deeper analysis has been conducted. In this study we proved that AAs along the backbone play an important role in dividing certain group of enzymes to different classes. Different PhaC synthases classes even related to one type of the enzymes (the polymerases) have differences in their end products, structures and substrate specificity. Why enzymes performing the same functions are different? Table 1 and Fig. 1 show that different classes have different AAs profiles even though they have minimum differences in their catalytic residues (Fig. 2) [13].

The AAs profile significantly affects SFS. One change in one amino acid could affect the fitting of the active site to its substrate. If such change becomes recognizable for the enzyme final products or the substrates, similar enzymes will be classified in one class. Changing the catalytic residue from one class to another one did not completely change one class to another [13]. PhaC synthases and proteases have been investigated in this study to answer this question. The AAs profile of each class of PhaC synthases and/or proteases have been compared with each other collectively or each type alone. The data in Table 1 prove that both of PhaC synthases and proteases are different in their AAs profile either between their different classes or between both of them. The differences between various classes could not be detected in case of investigating differences between 3D models as shown in Fig. 3 (a - e). Even though the models different but they aligned perfectly to each other (Fig. 3e). The differences on AAs affect on each class activity [4-6]. Phylogenic tree as in Fig. 1 perfectly group each class separately from the other except Class IV which is clustered with Class III (Fig. 1). This proves that the AAs similarity within one class is higher than within the overall classes which further support our hypothesis. Each class is unique in its SFS not only on its catalytic residue but on its constituent AAs.

Understanding the enzymes produced by a particular microbial strain is a key future for successful design of medium constituents especially if the process needs more than one enzyme such as the phenol degradation enzymes. Some substrates could inhibit the production of particular products. Using the available data present in the protein databases will lead to reduction in time for the production and optimization of particular products.

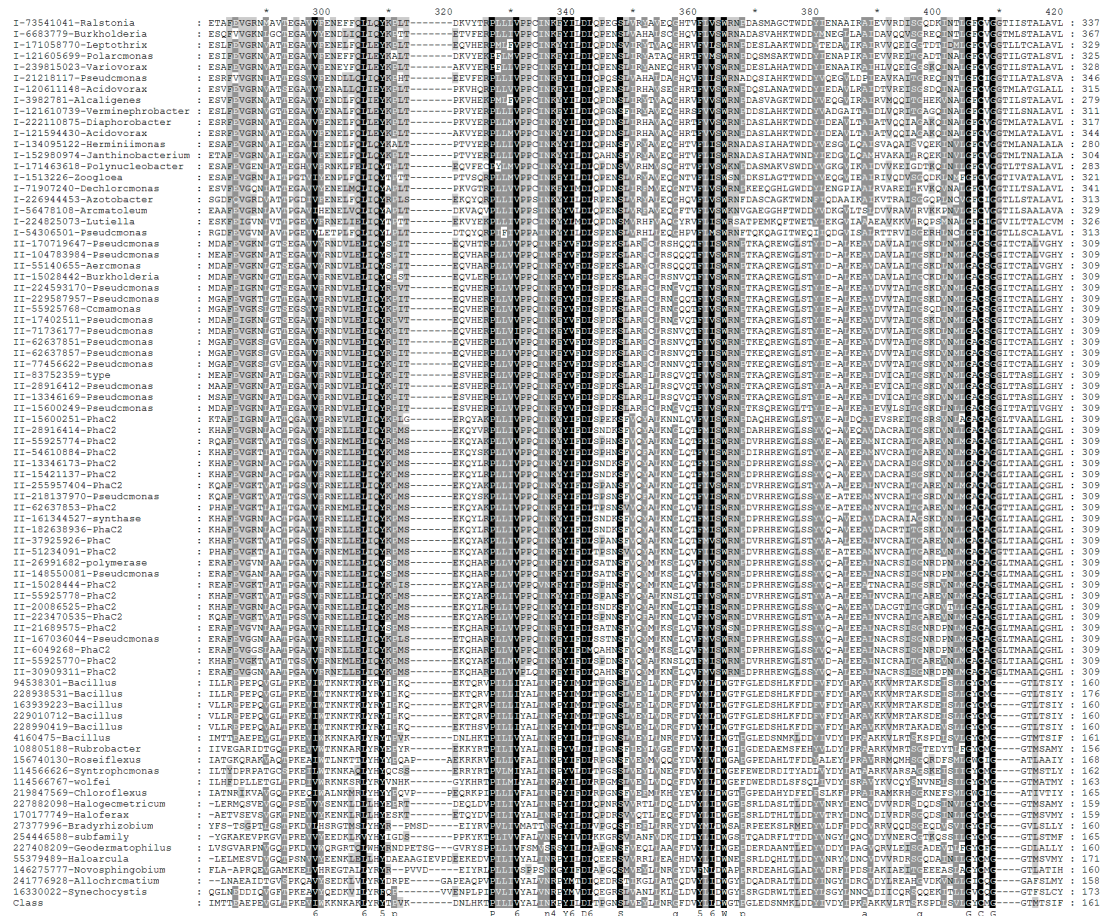


Figure 2: Partial multiple alignment of primary structure of 81 PhaCs different classes. Last line represent Class IV (*Bacillus megaterium*), shaded and less shaded represent conserved and highly conserved regions respectively

Microbes which are able to produce lipases when cultivated on castor oil as a substrate have been used in our model to establish the link between the available sequences in the databases and the possibility of designing a successful production medium. Whether or not it is possible to use the protein database in our experimental design, we establish a simple investigation method. In our experiment we use castor oil to produce lipases. We are looking for a strain that could survive in the presence of phenol (which is present in the castor oil crude extracts) [14]. The first two enzymes in the phenol degradation pathway are phenol 2-monooxygenase and catechol 1,2-dioxygenase whereas lipase is the enzyme which will degrade castor oil. Six strains from our lab stock namely *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Bacillus pumilus*, *Bacillus thuringiensis*, *Bacillus licheniformis* and *Geobacillus stearothermophilus* were screened for the three enzymes. Only *P. aeruginosa* proved to have the three enzymes as the databases tell us [data not shown]. This tool can solve many technical problems and can save a lot of time. As examples, Amara and Salem [10] show lipases production from a pharmaceutical grade castor oil and a recent study by Amara [11] showed the possibility to produce lipases, polyhydroxybutyrate and mesophilic proteases simultaneously using experimental design [11]. In fact one should observe that lipases is produced by *P. aeruginosa* outside the cells and another extra product(s) can be produced in

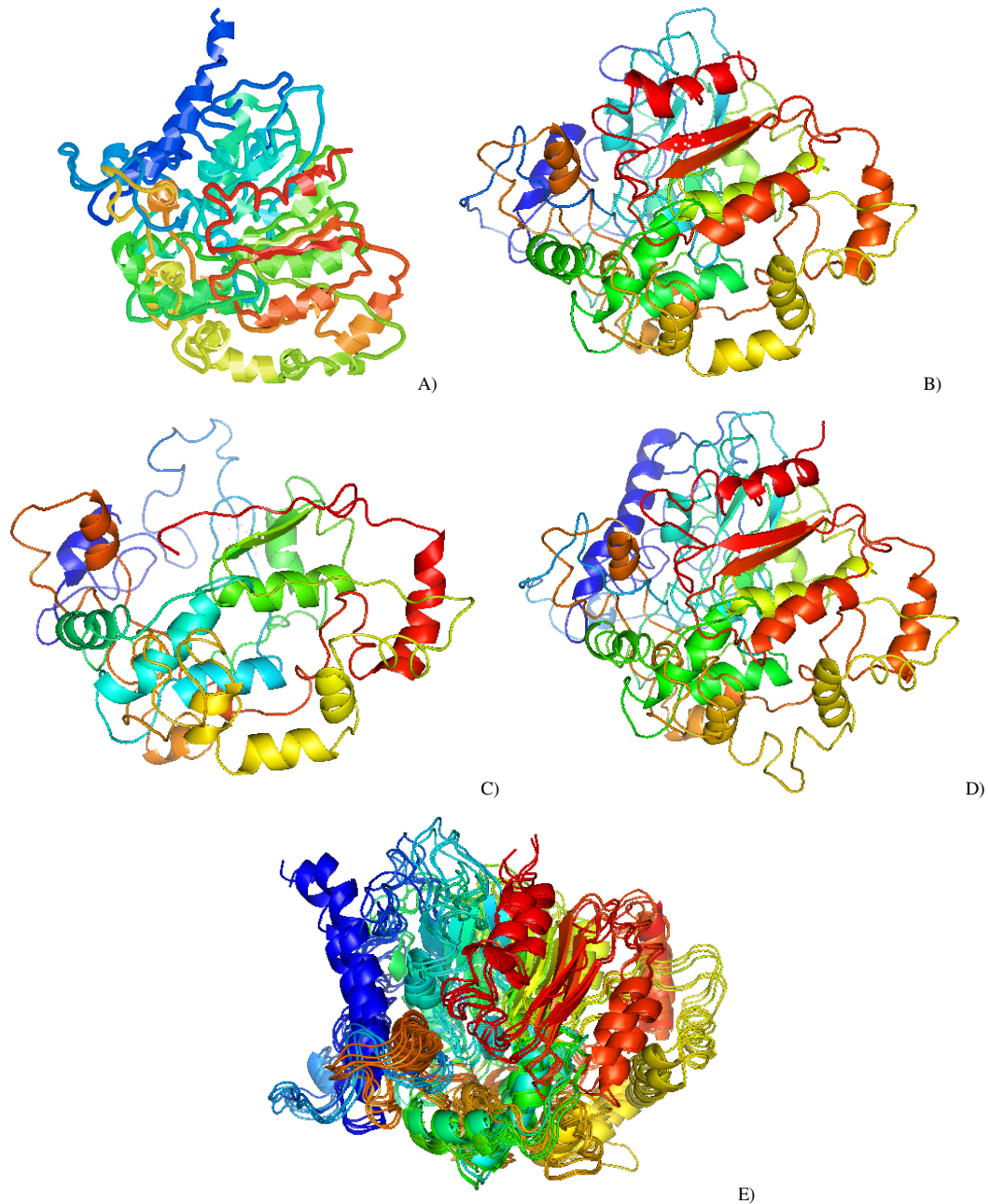


Fig. 3 A) *Ralstonia eutropha* model Class I; B) *Pseudomonas putida* model Class II; C) *Bacillus thuringiensis* model Class III; D) *Bacillus megaterium* model Class IV; Alignment of A, B, C and D models.

the same process such as PHAs which can be produced inside the cells as described by Amara [11]. However, both studies [10, 11] could face technical problem when oil source such as crude castor oil contain or is contaminated with traces amount of phenols. Crude castor oil contaminated naturally by phenols and other types of the phenolic compounds [14]. *P. aeruginosa* produces phenol degrading enzymes and lipases as expected from searching and analyzing sequences of the protein database, and could tolerate 1% phenol. The apparent lipase(s) activity is about 1070 Units/ml. It is logical to get the benefit from the available protein and DNA databases instead of running many experiments which will give

the same results, and in doing so will reduce the research time, cost and will give fast results. The strategies involved in our study successfully in solving such a problem.

The strategies included in this study are recommended for study different proteins which are able to do the same function but have specific differences like PhaC synthases and for establishing a new trend in experimental design. If a researcher get a new sequence or a new clone representing one of the PhaC synthases, using the database to determine where it will be located (in each class) or to determine the nearest similar enzyme; this will give the researcher a valuable information about how to design a medium for PHAs production.

4. CONCLUSION

This study addresses two main questions. The first deals with AAs constituent in enzyme other than the catalytic residues which appear to have critical function in catalysis, substrate specificity and function (SFS). The second deals with what benefits we can obtain from the available in the different biological database in experimental design. This study has answered both questions. PhaC synthases and proteases prove clearly the role of the AAs in SFS. The four classes of PhaC synthases have few differences in the catalytic amino acids which when changed did not turn one class to another one as proved by Amara and Rehm [13]. Even they have similar 3D structures (Fig. 3) they are classified into four classes based on their AAs (Table 1 and in Fig. 1). This proves clearly that the overall AAs profiles has significant role in SFS. Proteases give the same results as in Table 1. To answer the second question different strains have been used to investigate the presence of the phenol degrading enzymes as well as the lipases from all the strain have been investigated against the different databases and only *P. aeruginosa* prove to be suitable for the three essential enzymes (lipases and two for phenol degradation enzymes). The results prove that databases can shorten and improve the experimental design.

REFERENCES

- [1] A.A. Amara, M.S. Matar, E. Serour, "Comparative study on the amino acids constituents of PhaC synthases Class I, II, III and IV" *EMBO-Conference-“Comparative genomics of Eukaryotic microorganisms*. 2009. 59.
- [2] E. Serour, A. A. Amara, M.S. Matar, "Comparative study between thermophilic and mesophilic lipases and proteases". *EMBO-Conference-“Comparative genomics of Eukaryotic microorganisms*. 2009. 77.
- [3] A. A. Ammara, "Polyhydroxyalkanoates: From Basic Research And Molecular Biology To Application". *IIUM Engineering Journal* 9.1 (2008): 37-73.
- [4] B.H.A. Rehm, R.V. Antonio, P. Spiekermann, Al.Fattah. Amara , A. Steinbüchel, "Molecular characterization of the poly (3-hydroxybutyrate) (PHB) synthase from *Ralstonia eutropha*: in vitro evolution, site-specific mutagenesis and development of a PHB synthase protein model". *." Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1594.1 (2002): 178-90.
- [5] A.A. Amara, B.H.A. Rehm, A. Steinbüchel, "Biopolymer overproduction by new mutants using simple methods for selection". *DAAD-Bioforum-Berlin-Grenzenlos Forschen''*, *DAAD, Biotechnologische Methoden* (2001): 231-39.
- [6] A.A. Amara, A. Steinbüchel, B.H.A. Rehm, "In vivo evolution of the *Aeromonas punctata* polyhydroxyalkanoate (PHA) synthase: Isolation and characterization of modified PHA synthases with enhanced activity".. *Applied microbiology and biotechnology* 59.4-5 (2002): 477-82.
- [7] S. Taguchi, A. Maehara, K. Takase, M. Nakahara, H. Nakamura, Y. Doi, "Analysis of mutational effects of a polyhydroxybutyrate (PHB) polymerase on bacterial PHB

- accumulation using an in vivo assay system".. ." *FEMS Microbiology Letters* 198.1 (2001): 65-71.
- [8] S. Taguchi, H. Nakamura, T. Hiraishi, I. Yamato, Y. Doi, "In vitro evolution of a polyhydroxybutyrate synthase by intragenic suppression-type mutagenesis". ." *Journal Of Biochemistry* 131.6 (2002): 801-06.
- [9] S. Mater, A.A. Amara, E. Serour, "Study the possibility of lipases overproduction using media constituents designed based on genomic comparative study". *Embo-Conference-Comparative Genomics Of Eukaryotic Microorganisms*. 2009. 78.
- [10] A.A. Amara, Soheir.R. Salem, "Degradation of Castor Oil and Lipase Production by *Pseudomonas aeruginosa*". ." *American-Eurasian Journal of Agricultural & Environmental Sciences* 5.4 (2009): 556-63.
- [11] A.A. Amara "Experimental design for simultaneous production of PHB, Proteases and lipases". *IIUM Engineering Journal* 12.2 (2011): 155-84.
- [12] Frobisher, Martin, "Fundamentals of microbiology, An introduction to the microorganisms with special reference to the procaryons". WB Saunders (1968).
- [13] A. A. Amara, and Rehm, B. H. "Replacement of the catalytic nucleophile cysteine-296 by serine in class II polyhydroxyalkanoate synthase from *Pseudomonas aeruginosa*-mediated synthesis of a new polyester: identification of catalytic residues". *Biochemistry Journal* 1.374 (2003): 413-21.
- [14] M.G. Godoya, M.L.E. Gutarraa, F.M. Maciel, S.P. Felixb, J.V. Bevilaquac, O.L.T. Machadob, D.M.G. Freirea, "Use of a low-cost methodology for biodegradation of castor bean waste and lipase production". *Enzyme and Microbial Technology* 44 (2009): 317-22.

Part of this study has been represented in posters and abstracts in the EMBO conference "Comparative genomics of Eukaryotic microorganisms (2009) - Spain