UNIVERSITY OF BRASÍLIA

POSGRADUATE PROGRAM IN BIOTECHNOLOGY AND BIODIVERSITY

PAULA ISTVAN

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF NOVEL LIPOLYTIC ENZYMES FROM A BRAZILIAN CERRADO SOIL METAGENOMIC LIBRARY

Brasília DF

2018

PAULA ISTVAN

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF NOVEL LIPOLYTIC ENZYMES FROM A BRAZILIAN CERRADO SOIL METAGENOMIC LIBRARY

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy (PhD) of Biotechnology and Biodiversity of University of Brasilia

Supervisor: Prof. Dr. Ricardo Henrique Kruger

Brasília, DF

2018

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF NOVEL LIPOLYTIC ENZYMES FROM A BRAZILIAN CERRADO SOIL METAGENOMIC LIBRARY

Brasília, June 15, 2018

_____

Prof. Dr. RICARDO HENRIQUE KRUGER

Supervisor

_____

Prof. Dr. JANICE LISBOA DE MARCO

University of Brasilia

_____

Prof. Dr. LIDIA MARIA PEPE DE MORAES

University of Brasilia

_____

Prof. Dr. MARCELO RAMADA

Catolic University of Brasilia

ACKNOWLEDGEMENTS

I wish to express my grateful thanks to my supervisor Prof. Dr. Ricardo Henrique Kruger for giving me the chance to work with him and I would also like to thank him for his supervision, encouragement, criticism and patience to guide me during my thesis studies.

I would like to thank my colleges from the Laboratory of Enzymology for valuable comments and help during my thesis studies and for sharing their experiences with me. My special thanks goes to Amanda who works with me in the Biophysical Laboratory for her sincere help offers, friendships and moral support as well to his colleges from there.

I want to express my grateful thanks to Clovis, Enrique and Antonio for being with me through all the times during this study for their motivation, support and endless love. Finally, I express my deep gratitude and appreciation to my family members for their unconditional love, belief, support and enormous encouragement.

RESUMO

O solo do Cerrado tem sido foco de poucos estudos, especificamente sua diversidade microbiana. O solo é um habitat com alta diversidade de microrganismos e, portanto, pode ser fonte para a prospecção de enzimas industriais. A identificação de novos genes pela abordagem metagenômica levou à descoberta de novas atividades enzimáticas. As enzimas lipoliticas são produzidas principalmente por microorganismos, frequentemente bactérias, e desempenham um papel vital em iniciativas comerciais. A fim de isolar genes putativos de lipase, foi realizado um *screening* com tributirina em uma biblioteca metagenômica de solo de Cerrado. De 6.720 clones avaliados contendo um inserto de mediana de 8 Kb, foram isolados três com atividade lipolítica e dezoito seqüências codificantes foram preditas. LipX e lipY se agruparam a enzimas da família IV lipolítica, relacionados com lipase de *Pseudomonas sp.* (AAC38151), *Cupriavidus necator* (AAC 41424) e *Moraxella sp.* (CAA37862). LipW apresentou semelhanças estruturais com uma enzima lipolítica de *Pseudomonas fluorescens* e 28% de identidade de sequência, no mesmo ramo de quatro lipases / esterases da bactéria *Moraxellaceae*. Além disso, a tríade catalítica Ser-Asp-His foi localizada no motivo G-X-S-M-G-G, similar a outras lipases da família V. A purificação da lipX revelou uma proteína recombinante de 35 kDa, e da lipY revelou uma proteína recombinante de 32 kDa como previsto pelos dados de sequenciamento. O peso molecular estimado do LipW foi de 29 kDa, os parâmetros moleculares puderam ser associados ao monômero do lipW em correspondência com o seu MW teórico, como também se observou no gel de SDS-PAGE. A atividade enzimática foi observada de pH 3,0 a 10,0, com atividade ótima em pH 7,5,0-9,5 para lipX, pH 8,0-9,5 para lipY e em pH 9,0-9,5 para lipW. Dentro da faixa de temperatura de 25°C-95°C, a atividade máxima foi observada a 55°C para lipX, 60°C para lipY e 40°C para lipW. Os resultados da comparação entre o lipX com o algoritmo BLAST e o banco de dados de proteínas (PDB) exibiu 41% de identidade com mutante G84S EST2 de *Alicyclobacillus acidocaldarius* [PDB 2HM7] e dobra α/βhydrolase da família sensível ao hormônio IV Esterase/Lipase. Com os resíduos catalíticos: Ser156, Asp259 e His289 localizados entre o domínio α/β e o domínio helicoidal. O Ser156 está localizado no motivo de consenso GDSAGG, e também foi visto no alinhamento de sequência de lipX com outros membros da família IV. LipY tem uma tríade catalítica formada por Ser143, Glu237 e His267. LipY também contém um domínio CAP (Met1–Val45) além do domínio catalítico (Gln46–Arg306). O Ser143 catalítico está localizado no motivo GXSAG semi-conservado onde, em vez de Asp (GDSAG), o mais comum, há outro resíduo de carga negativa, Glu (GESAG). LipX e lipY contêm motivo HGG conservado, que está envolvido na formação do canal oxianionico. Juntos, os espectros de CD UV distante e a estrutura do modelo 3D baseada na modelagem por homologia indicam que a estrutura secundária lipX, Y e W consistia em α-hélice (~45%) e estruturas de folha-β (~15%), o que é consistente com a estrutura secundária predominantemente α-hélice da família α/β-hidrolase de esterases. A combinação de propriedades estruturais e funcionais demonstra que a abordagem metagenômica de ambientes nativos é adequada para descobrir novas enzimas lipolíticas com aplicação biotecnológica.

Palavras-chave: Solo, Cerrado, Enzimas lipoliticas, Metagenoma.

ABSTRACT

The Brazilian Cerrado soil has been focus of very few studies about its microbial diversity. Soil is a habitat with high diversity of microorganisms, thus can be a source for the prospection of industrial enzymes. The identification of new genes through the metagenomic approach has led to the discovery of novel enzyme activities. Lipases are mainly produced by microbes, most frequently bacterial, and play a vital role in commercial ventures. In order to isolate putative lipase genes, Cerrado soil metagenomic library was screened with tributyrin. From 6,720 clones evaluated with 8 Kb DNA size, three with lipolytic activity were isolated, and eighteen coding sequences were predicted. LipX and lipY clustered with lipolytic enzymes from family IV, closely related to a lipase from *Pseudomonas sp.* (accession number AAC38151), *Cupriavidus necator* (accession number AAC 41424) and *Moraxella sp*. (accession number CAA37862). LipW showed structural similarities with a *Pseudomonas fluorescens* lipolytic enzyme, and 28% sequence identity in the same branch as four lipases/esterases from Moraxellaceae bacteria. In addition, the Ser-Asp-His catalytic triad was localized in the G-X-S-M-G-G motif, similar to other family V lipases. The purification of lipX revealed a 35-kDa recombinant protein and lipY revealed a 32-kDa recombinant protein as predicted by the sequence data. Estimated molecular weight of LipW was 29 kDa, molecular parameters could be associated to the lipW monomer in correspondence with their theoretical MW, as also observed in the SDS-PAGE gel. Enzyme activity was observed from pH 3.0–10.0, with optimal activity at pH 7.5.0–9.5 for lipX, pH 8.0–9.5 for lipY and at pH 9.0–9.5 for lipW. Within the temperature range of 25°C–95°C, maximum activity was observed at 55°C for lipX, 60°C for lipY and 40°C for lipW. The results from the comparison of the lipX with BLAST algorithm against protein data bank (PDB) exhibited 41% identity with G84S EST2 mutant from *Alicyclobacillus acidocaldarius* [PDB 2HM7] and α/βhydrolase fold from the IV Esterase/Lipase hormone sensitive Family. With the catalytic residues: Ser156, Asp259 and His289 located between α/β domain and the helical domain. The Ser 156 is located in the consensus motif GDSAGG, as also showed in the sequence alignment of lipX with other members of related Family IV. LipY has a catalytic triad formed by Ser143, Glu237, and His267. LipY also contains a CAP domain (Met1–Val45) besides catalytic domain (Gln46–Arg306). The catalytic Ser143 is located in the semi-conserved GXSAG motif where instead of Asp (GDSAG), the most common; there is another negative-charged residue, Glu (GESAG). LipX and lipY contain a conserved HGG motif, which is involved in the formation of the oxyanion hole. Together, the far-UV CD spectra and 3D model structure based on homology modeling indicate that lipX,Y and W secondary structure consisted of α-helix (~45%) and β-sheet (~15%) structures, which is consistent with the predominantly α-helix secondary structure of the α/β-hydrolase family of esterases. The combination of structural and functional proprieties demonstrates that metagenomic approach of native environments is suitable for discovering novel lipolytic enzyme with biotechnological application.

Keywords: Soil, Cerrado, Lipolytic enzymes, Metagenome.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Å	Angstrom
LB	Luria-Bertani
UV	Ultraviolet
rpm	Revolutions per minute
PCR	Polymerase chain reaction
OD	Optic density
rRNA	Ribosomal ribonucleic acid
rDNA	Ribosomal deoxyribonucleic acid
ng	Nanogram
µg	Microgram
µl	Microliter
ml	Milliliter
µM	Micromolar
mM	Millimolar
IPTG	Isopropyl-thio-β-D-galactopyranoside
BLASTp	Basic local alignment search tool for protein
min	Minute
sec	Second
w/v	Weight per volume
v/v	Volume per volume
µm	Micrometer
nm	Nanometer
sp.	Species (singular)
bp	Base pair
kb	Kilobase pairs
kDa	Kilodalton
Ω	ohm

TABLE OF CONTENTS

Biocatalysts were first described at the end of the 18th century, with studies of meat digestion by stomach secretions. However, it was from the 19th century that the nature of these catalysts began to understand, with studies regarding starch conversion by sugars, saliva, and plant extracts. In Germany in 1877 Wilhelm Kühnen introduces the term "enzyme". Enzymes are capable of performing many different reactions, can be produced on a large scale, and operate at different temperatures and pH (Salis et al., 2007). These properties have captured the attention of scientists and engineers for the use of enzymes as industrial catalysts, used extensively across a wide range of applications as food industry, organic chemistry, detergency and cleaning, paper industry, management of waste and toxic compounds, as components of biosensors, in biofuels production, in leather processing, in hard-surface cleaning, in single-cell protein production, and in the synthesis of polymers, biodegradable plastics, lubricants or cosmetics (Choudhury and Bhunia, 2015a).

Lipolytic enzymes are ubiquitous enzymes. In eukaryotes, they may be find within organelles, such as lysosomes, or they may be found in spaces outside cells, which play an important role in the metabolism, absorption and transport of lipids. In lower eukaryotes and prokaryotes, the lipases may be intracellular or secreted for the purpose of degrading lipid substrates present in the environment One important aspect of lipolytic enzymes is the unique physico-chemical character of the reactions they catalyse at lipid-water interfaces. They were previously defined in kinetic terms, based on the “interfacial activation” phenomenon, in terms of the increase in the activity which occurs when a partially water-soluble substrate becomes water-insoluble (Khan et al., 2017).

Several studies are focused to find a different source that can provide microorganisms accessing uncultivable organisms and enzymes with a great biotechnological potential and industrial usage.(Berini et al., 2017). From all sources studied, the soil represents the main source of the majority biocatalysts, antibiotics and secondary metabolites of importance in industry, medicine and agriculture and only a small fraction of total microbial species in soil has been characterized by cultivation-based methods (Torsvik et al., 1990). Bioprospection studies of microbial diversity in soil revealed that the *Proteobacteria, Acidobacteria, Actinobacterium, Firmicutes, and*

Verrucomicrobia are abundant (Delmont et al., 2011), as well as several of uncultured bacteria, which are difficult to access.

Nevertheless, regarding the uncultured bacteria, it is worth to mention that metagenomics technique made the access to such kind of microorganism possible and it can be understood as a technique in which a portion of the DNA is extracted and inserted in a culturable surrogate host (Gu et al., 2015). Considering metagenomics from a soil, a number of genes encoding novel enzymes, and, as in this work, lipolytic enzymes, biocatalysts for biotechnological application have been identified and characterized beside this novel families of esterase/lipase enzymes settle (Lee et al., 2006).

For the enzymatic technology, a challenge to be overcome is to optimize the production of these enzymes on a large scale. Currently, some expression systems are being studied extensively, among them the classics like Gram-negative bacterium *Escherichia coli*, as well as alternative systems such as the methylotrophic yeast *Komagataella pastoris* and filamentous fungi (Valero, 2012). These latter systems have gained wide acceptance as host organisms in the production of heterologous or homologous proteins of industrial interest, due to the efficiency of the process and because they are more suitable to express new lipolytic genes from eukaryotic origin using metagenomic aprouch (Berini et al., 2017; dos Santos et al., 2017; Stroobants et al., 2015). The present work addresses those issues by the functional metagenomic screening of a metagenomic library from Brazilian Cerrado soil samples, the cloning, overexpression, functional and structural characterization analysis of tree novel lipolytic enzymes lipX, lipY and lipW.

Global market of enzymes

Modern biotechnology encompasses the application of industrial enzymes with the possibility of finding new biomolecules using improved processes, efficient microbial producing species, the orientation of enzymes to specific reactions in metabolic engineering and metagenomics (Borrelli and Trono, 2015). Most of the enzymes used on industrial scale are currently produced by genetically modified organisms and are applied in various sectors: pharmaceutical, detergents, textile finishing, pulp and paper processing, food processing, chemical production.

The global market for industrial enzymes has been growing significantly year after year. In 2014, the total market reached approximately USD 4.6 billion, while in 2016 it has surpassed the mark of USD 5 billion, an increase of 8,6% in the period. For 2021, the expectation is for the market to reach USD 6,3 billion, with an annual growth rate (CAGR) of 4.7% (Dewan, 2017). Companies that have successfully developed commercial enzymes build up the same skills: (i) screening for new enzymes; (ii) fermentation; (iii) large scale purification; and (iv) formulation of commercial enzymes. Some of these companies in the Enzymatic Industry are: BASF SE (Germany), E.I. DuPont de Nemours & Co (U.S.), Koninklijke DSM N.V (Netherlands) and Novozymes A/S (Denmark).

Nevertheless, the main challenges for the adoption of biocatalysts as a regular industrial practice are: (i) costs, since the biocatalysis process is more expensive than ordinary chemical methods; and (ii) although enzymes can show a better performance than chemical methods in their native habitat, most of them are not suited for commercial use, since conditions are different in industrial and natural environments.

Despite all such challenges, enzymes as a solid established biotechnology product are used mainly in the food sector (31% of the global sale of enzyme), with a significant potential growth in the development countries. It is important to highlight that there are sectors in which biocatalysts can be a resourceful tool such as wastewater treatment, and paper and pulp sectors (Sarrouh et al., 2012).

Lipolytic enzymes

Lipids are essential to all living systems. They are relevant sources of energy, performing structural roles in the membrane and involved in the cellular signaling. To perform these functions, they require lipolytic enzymes during metabolism (Facchin et al., 2013).

Lipolytic enzymes were discovered in 1856 by Claude Bernard while studing the role of pancreas in fat digestion (Arpigny and Jaeger, 1999). Since then, several different enzymes capable of acting over lipids have been identified and isolated from bacteria, fungi, plants and animals. Due to their diversity and catalytic properties, lipase has a prominent spot as a relevant industrial biocatalyst.

Lipolytic enzymes belong to the group of the carboxyl ester hydrolases, which are represented mainly by esterase (E.C.3.1.1.1) and "real" lipase or triacylglycerol hydrolases (E.C.3.1.1.3) that differ in some biochemical aspect. Lipases catalyze the hydrolysis of the ester connection with triacyclglycerols in a chain with more than ten carbons; and esterases, which catalyze the hydrolysis of triacyclglycerols constituted by oily acids with less than ten carbons. However, other enzymes can hydrolyze acyl glycerol such as the cutinases and phospholipases, considered by some authors as lipolytic enzymes (Fojan et al., 2000).

Lipase acts in the aqueous/organic interface, catalyzing the hydrolysis of the ester-carboxylic connections and releasing acids and organic alcohols. However, in the aqueous-restricted means, the inverted reaction (esterification) or various reactions of transesterification can also occur. Therefore, lipases are capable of catalyzing hydrolysis reactions as much as synthesis in triglyceride and fatty. This diversity of reactions presented by the lipase has a great significance to the industry, especially to the production of chemical compounds, pharmaceutical cosmetics, detergents and biodiesel. Lipases can be highly specific regarding the compound they hydrolyze/synthetize.

This specificity is controlled by the molecular properties of the enzyme, unsaturation size and number of the substrate carbonic chains and by factors that affect the enzyme-substrate connection. In this context, steric obstructions, as well as the hydrophobic interactions, are crucial to the enzymatic specificity (Fojan et al., 2000).

Biological sources of lipolytic enzymes

The lipolytic enzymes can be found in plants, animals and micro-organisms. Plant lipases are found mainly in tissues, as, for example, in oily seeds, cereal seeds and tree barks. Thus, they have potential for commercial application in organic synthesis, food, detergent and pharmacology (Choudhury and Bhunia, 2015b). Alongside with plant lipases, there are the animal lipases, which include the liver lipase and pancreatic lipases and the ones sensitive to hormone and the ones stimulated by the bile. Those with pancreatic origin are used in the acceleration of the cheese maturation and in the creation of flavor, as well as in the lipolysis of butter, grease and cream (El-Hofi et al., 2011).

Due to their catalytic properties, the microbial lipases have a prominent position among the industrially used biocatalysts. The first bacterial lipases acknowledged as members of the enzyme superfamily were homologous to the ones from animal origin. The most important bacterial genera that have been studied for the production of lipase are *Pseudomonas, Bacillu*s e *Streptomyces*, followed by *Burkholderia, Chromobacterium, Achromobacter, Alcaligenes* e *Arthrobacter*. From all these, a number of bacterial lipases are used in industrial processes, such as the production of biodiesel, particularly lipases of *Pseudomonas fluorescens, Burkholderia cepacia and Bacillus thermocatenulatus* (Hwang et al., 2014) while the lipases from *Pseudomonas mendocina, Pseudomonas alcaligenes*, *Pseudomonas cepacia glumae* are used as detergent additives (Gupta et al., 2004). Among the relevant kind of yeasts and fungi studied for the production of lipases, one can include the *Candida*, *Yarrowia*, *Aspergillus, Penicillium, Rhizopus*, *Rhizomucor e Thermomyces* (Hwang et al., 2014).

Classification of the lipolytic enzymes

Due to the difficulty to classify the lipases based on the mechanism of reaction, Arpigny e Jaeger (1999) proposed a classification system based on the similarity between the sequence of several enzymes and the lipase from *Pseudomonas aeruginosa* (100%). The sequences of amino acids and nucleotides were obtained from the data stored in the NCBI and the comparison between them relied on the BLAST 2.0. program (Altschul et al., 1997).

The classification suggested 8 families of lipolytic enzymes, being that families II and VIII encompass esterases. Since then, this classification has been largely used and integrated to various database servers. There are currently 16 families of lipolytic enzymes.

Family I consists in true lipases defined as enzymes which show maximal activity towards insoluble long-chain triglycerides. These enzymes are subdivided in:

Sub-family 1: Lipases that present sequence similar to *Pseudomonas aeruginosa (P. aeruginosa)* lipase and whose mass is usually between 30-32kDa. These lipases require chaperone proteins for expression. All members have two conserved aspartic acid residues for controlling Ca2+-binding site for catalytic activity. Besides, most of them contain an intramolecular disulphide bridge for stabilizing protein folding. These features are found near the catalytic histidine and aspartic acid residues.

Sub-family 2: Lipases that show sequence similarity to *Burkholderia glumae (B. glumae)* lipase and whose mass exceeds 32 kDa (with two extra β-strands when compared to sub-family 1). Like sub-family 1, these lipases need chaperone for expression, as well as the two aspartic acid residues and two cysteine residues forming an internal disulphide bridge.

Sub-family 3: Lipases from two separate species – *Pseudomonas fluorescens (P. fluorescens) and Serratia marcescens (S. marcescens*). Their molecular mass is greater than sub-families 1 and 2 (50-65kDa), with no N-terminal signal peptide or disulphide bridge. These lipases are exported from host through type I secretion pathway using a C-terminal signaling domain.

Sub-family 4: These are the smallest lipases, with a molecular mass of less than 20kDa. Several contain the pentapeptide sequence AXSXG instead of GXSXG. They do not require Ca^{2+} ion for catalytic activity and bear no cysteine residues.

Sub-family 5: All lipases in this family are from gram-positive prokaryotes, with molecular mass ~46 kDa due to a unique insertion within the α/β hydrolase fold required for zinc-binding, theorized to be behind thermal stability.

Sub-family 6: Lipases in this family start as preproproteins ~75 kDa in mass due to ~200 amino acid N-terminal domain used as a translocator signal through the cell membrane. This precursor is cleaved in the extracellular medium through specific protease activity, leaving a lipolytic domain ~46kDa. Some members have shown phospholipid as well as lipolytic activity.

Sub-family 7: Central region primary structure (i.e. from residues 50-150) shows significant similarity to sub-family 2. Members of this family have been found to act on a wide range of substrates, both tri- and mono-glyceride in origin, of varying fatty acid chain lengths.

Sub-family 8: Recently identified as a novel lipase – Lip1 – from *Pseudoalteromonas haloplanktis (P. haloplanktis).* The primary structure of this 51 kDa lipase shows little relation to any previously identified esterase's. They theoretically lack lid structures and Ca^{2+} pockets and the active site is not the expected GXSXG or equivalent pentapeptide, but instead LGG(F/L/Y) STG heptapeptide (Arpigny and Jaeger, 1999)

Family II is encompassed by GDSL/SGNH esterases, lacking the pentapeptide GXSXG sequence and instead containing GDSL/SGNH tetrapeptide at the N-terminus. Family II esterases have five sections/blocks bearing conserved amino acid residues, with the first block containing the GDSL motif (Lenfant et al., 2013). As with GXSXG esterases, it is the serine residue in these GDSL/SGNH esterases which performs a nucleophilic attack on the substrate at the binding site of the esterase.

Family III: Members of this family show sequence similarity (~20%) to human platelet activating-factor acetylhydrolase (PAF-AH). However, family III esterase's do not have a lid, unlike PAF-AH. The first members were the extracellular esterase's with a molecular mass of 32-35 kDa from the species *Streptomyces and Moraxella*. Later additions included *Acidovorax and Thermobifida* hydrolases, both of which are capable of degrading polyesters.

Family IV: A series of esterase's from distantly related prokaryotes that show sequence similarity to mammalian HSL (hormone-sensitive lipase). Members of this family have three sequence domains with conserved motifs, with domains II and III containing the esterase catalytic triad. Domain I contains a consensus sequence for stabilizing the oxyanion hole and promoting catalysis. Some members can also have a 'cap' which covers the active site, regulating the hydrolytic activity, but this does not configure them as actual lipases. Unlike HSL, these enzymes show greater activity on soluble, short-chain esters and expected kinetics upon substrates.

Family V: These esterases show homology to other, non-lipolytic hydrolases showing the α/β hydrolase fold and catalytic triad (Lenfant et al., 2013) Like Family IV, they have three conserved sequence blocks with the active site residues located in blocks II and III.

Family VI: Among the smallest of the esterases, with sizes ranging from 23-26kDa. Like in families IV and V, they have three conserved blocks. Family VI enzymes show ~40% sequence similarity to eukaryotic lysophospholipase.

Family VII: Members of this family have a mass ~55 kDa and four conserved sequence blocks. Amino acid sequence homologies to acetylcholine esterase's and intestine/liver carboxylesterases (30% identity and 40% similarity) respectively.

Family VIII: Unlike all other carboxylesterases, members of thi family do not have the α/β hydrolase fold yet show greater similarity to the β-lactamases and DD-peptidases. The catalytic serine is not found in a triad arrangement, but as part of a SXXK tetrapeptide as opposed to a GXSXG pentapeptide or a GDSL/SGNH tetrapeptide. Most members have a mass ~42 kDa. The serine in the SXXK tetrapeptide acts as in GXSXG or GDSL/SGNH esterase's (Biver and Vandenbol, 2013).

More esterases that have being discovered are not applicable to the current definitions of families/sub-families. The metagenomic studies have expanded this classification to include six new families, ranging from IX to XIV (Zarafeta et al., 2016). Some of the structural variability within the α / β hydrolase fold family is showed in Figure 1.

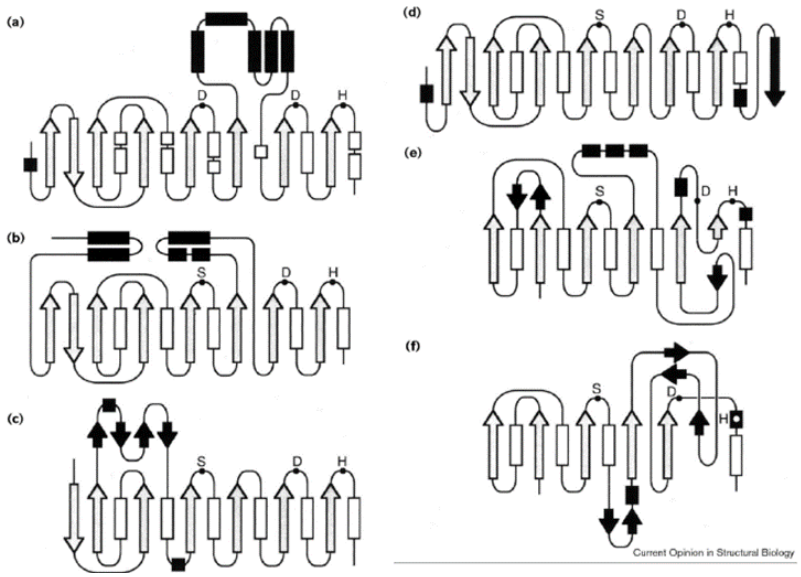


Figure 1 – Structural variability within the α / β hydrolase fold family.
α Helices and β strands belonging to the 'canonical' fold is represented by white rectangles and gray arrows, respectively. Helices are shown by squares. Secondary structure elements that deviate from the 'canonical' fold is represented in black. (a) Epoxide hydrolase from A. radiobacter AD1; (b) Brefeldin A esterase from B. subtilis; (c) Carboxylesterase from Pseudomonas fluorescens; (d) Lipase from S.

exfoliates; t. (e) Lipase from P. aeruginosa; (f) Acetylxylan esterase from P. purpurogenum (Nardini and Dijkstra, 1999).

Another classification allocates the lipases into three classes according to the oxoanionic hole: GX, GGGX, e Y. The oxoanionic hole is a pocket that promotes the stabilization of the intermediary residue charged negatively and generated during the ester connection in the hydrolysis. It consists of two residues, one always nucleophilic and the other with a changeable position. In the GX class of lipases, the hole position in the changeable residue C-terminal of X is next to the residue preserved glycine. In class GGGX, this residue is dislocated to a position related to the C-terminal and it is the glycine residue that is followed by a hydrophobic residue preserved X. In the Y class, the changeable residue is a tyrosine. Based on this classification and in the similarities of the amino acid sequences, the yeast and fungus lipases were grouped into five different subclasses, two in the GX class, two in the GGGX class and one in the Y class (Borrelli and Trono, 2015). The specificity of this substrate enzyme is directly correlated to the preference of the real lipase and esterase regarding the hydrophobicity and the physical state of their substrates (Figure 2).

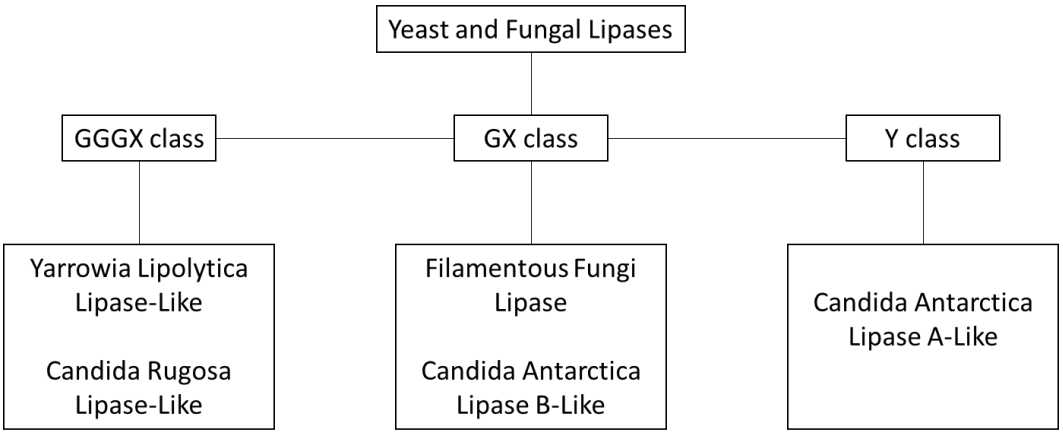


Figure 2 – Classification of yeast and fungal lipases.
(Adapted from (Borrelli and Trono, 2015)

Structure and catalytic mechanism of the lipases

The tertiary structure of lipases is characterized by a structural domain of the hydrolases of type α/β (Khan et al., 2017). The catalytic site of the lipases is composed by cyatalytic triad (G-X1-S-X2-G, where G=glycine; S=serine; X1= histidine e X2= glutamic acid and aspartic) (Figure 2). The components of the catalytic site are always in this order in

the amino-acid sequence, which is different from that observed for other proteins that carry the catalytic triad. In the lipases, the nucleophilic residue is always a serine, which is located in the so-called "nucleophilic elbow". The catalytic acidic residue is located after the β6 or β7 strand of the central β sheet, and it is hydrogen-bonded to the catalytic histidine that is located in a loop after the β8 strand of the α/β hydrolase fold (Figure 3). The active site of the lipase also contains a large, hydrophobic scissile FA binding site that accommodates the acyl chain of the ester linkage that is to be hydrolyzed.

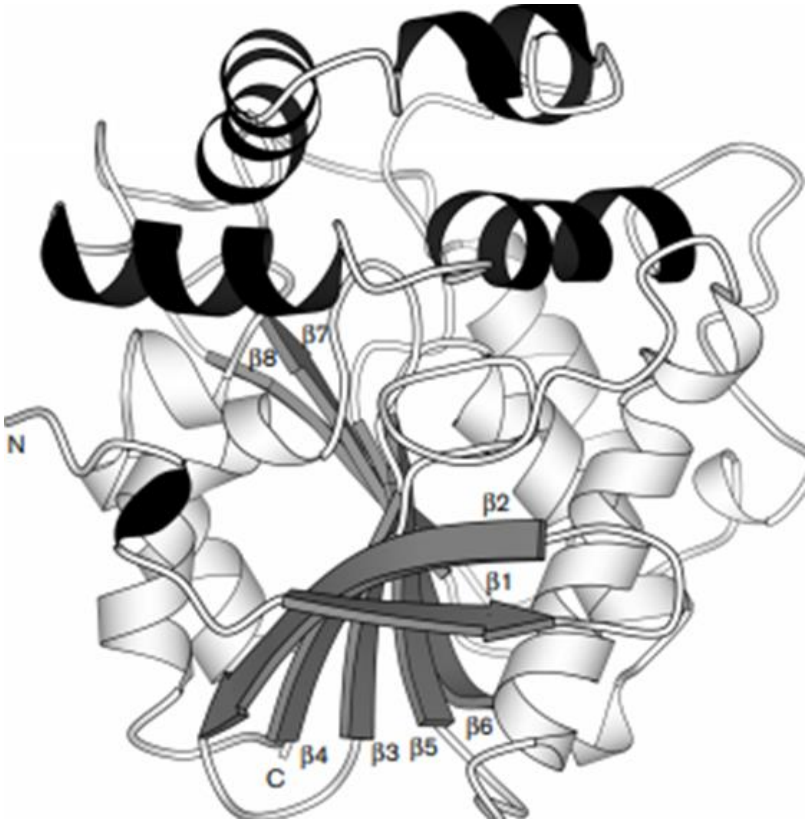


Figure 3 – Schematic representation of the structural r preserved in enzymes of the family α/β hydrolysis.*One β central leave, encompassing eight different β strips (β1-β8), connected to six α helix. (Adapted from (Nardini and Dijkstra, 1999)*

Most lipases are characterized by inclusion in their structure of a mobile domain, known as the "lid", which consists of a single helix, two helices, or a loop region. In the absence of a lipid–water interface, the lid covers the active site, while the presence of an interface results in a rearrangement of the conformation of the enzyme that displaces the lid, thus making the catalytic site accessible to the substrate and the solvent. This phenomenon was described by Holwerda et al (1936) and by Schonheyder & Volqvartz (1945) as a characteristic of the lipases that distinguishes them from the esterases. By

measuring the activity of the pancreatic lipase using the tricaproin as substrate, the authors realized that the hydrolysis was increased whenever the substrate concentration exceeded solubility limit. This behavior was called interfacial activation.

The lipase lid is an amphipathic structure and when closed its hydrophilic side is in contact with the aqueous solvent, while the hydrophobic side is facing the hydrophobic center of the protein. At the water-lipid interface, the lid moves, exposing the hydrophobic side, contributing to the formation of a larger hydrophobic surface and the attachment of the substrate at the same time as exposing the active site. The structure of the lid can be opened and closed and it is present on the active site in different lipase families, indicating that it would be involved in the catalytic mechanism of lipases (Cygler and Schrag, 1999) as showed in Figure 4.

Some lipases may be in dynamic equilibrium between the open and closed form, when in solution. In addition, the lid opening can be facilitated by organic solvents, probably by decreasing the dielectric constant of the medium (Anobom et al., 2014)

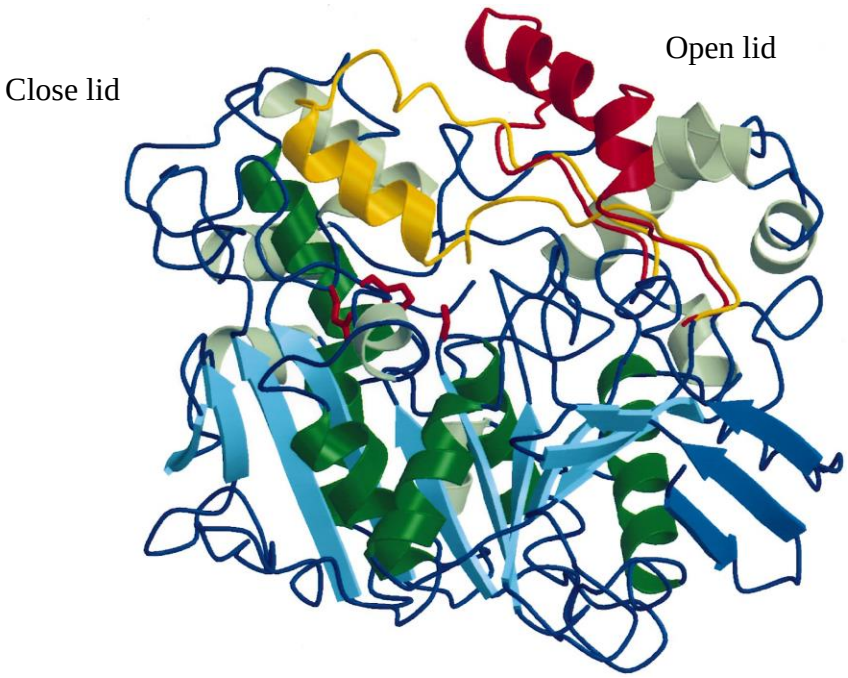


Figure 4 – Schematic representation of the structural open and closed states of the lid.
The closed conformation of the lid is yellow and the open conformation is red. (Cygler and Schrag, 1999).

Applications of lipolytic enzymes

Lipolytic enzymes constitute an important group of biotechnologically important enzymes due to the versatility of their properties and the large-scale production facility.

Among the possibilities of industrial applications of lipolytic enzymes are the use in the food industry together with other enzymes for the elaboration of bread, cheese, lipolysis of butter, fats, and cream, to produce aromas, emulgents and to produce flavors (Guerrand, 2017).

They are used in pharmaceutical and agrochemical industries for the modification or synthesis of antibiotics, anti-inflammatory compounds, pesticides, etc., and for the production of enantiopure compounds and the resolution of racemic mixtures (Kohli and Gupta, 2016)

The quantitative determination of triacylglycerol by lipolytic as a biosensor is of great importance in clinical diagnosis for the determination of triacylglycerol's. *Candida rugosa* lipase biosensor from *Candida rugosa* has been developed as a DNA (Choudhury and Bhunia, 2015a; Zehani et al., 2014). They can also be used to accelerate the degradation of fatty and polyurethane waste, contributing to the treatment of effluents and bioremediation (Guerrand, 2017; Shelat and Padalia, 2016; Singh et al., 2016).

An important application of lipolytic enzymes resistant to proteolysis and denaturation by surfactants is their use in laundry detergents along with proteases to improve the removal of lipid stains. They are also used in the synthesis of surfactants for soaps, shampoos, and dairy products These enzymes are biodegradable, they do not leave any harmful residues and risks for aquatic life. (De Godoy Daiha et al., 2015; Guerrand, 2017) They are used as well in production of biodiesel (Hwang et al., 2014).

Examples of some commercially available lipolytic enzymes are: Palatase® (Novozymes) from *Rhizomucor miehei*, used in the food industry; Chirazyme®L-8 (Boehringer Mannheim) and SP524 (Nova-nordisk) from *Geotrichum candidum*, used in oleo chemistry; Chirazyme®L5 (Boehringer Mannheim) and SP526 (Nova-Nordisk) from *C. antarctica*, used in organic chemistry; Lipase M "Amano" 10 (Amano), from *Mucor javanicus*, used in food processing and oleo chemistry, Lipolase® from *Thermomyces lanuginose* (Nova Nordisk), Lumafast® (de *Pseudomonas mendocina* (Genecor), used in detergent industry and Lipomax™ from *Pseudomonas alcaligenes* (Singh and Mukhopadhyay, 2012).

Metagenomic approach

The prospection of microbial enzymes imposes a relevant challenge, once most of the microorganisms found in a given environmental sample are non-cultivable yet (Delmont et al., 2011; Handelsman et al., 1998). In that sense, one of the strategies most likely to access the totality of micro-organisms in a given environmental sample and, therefore, find relevant biomolecules is the metagenomic approach.

In the metagenomic approach, the total DNA of micro-organisms found is isolated and cloned into specific vectors for the development of metagenomic libraries. Among such environmental samples available, there are several kinds of soil, water from lakes, seas, mud, rumen and drinkable water. It is known that 1g of soil can contain around 10 billion micro-organisms, encompassing thousands of different species (Torsvik et al., 1990). Yet, as abovementioned, most of these microorganisms are yet to be discovered due to their inability to multiply in culture commonly used in laboratories. Therefore, the metagenomic technique allows a sampling of these micro-organisms' genetic material to be much more substantial than the sampling obtainable from the analysis of cultivable micro-organisms alone, and enables, among other resources, the search for hydrolytic enzymes with industrial interest.

As shown in Figure 5, applications that use this approaches include the identification of genetic variants, description of diversity and abundance functional and taxonomic and the assembly of microbial genomes, this made by bioinformatic analysis. Still, the sequential metagenomics is less complex, since the steps of cloning and expression are not required. In the other hand, the functional screening is still a robust technique to identify new biocatalysts of interest from different environmental samples using molecular modeling and biodynamic analysis of the new molecules. In that sense, fragments of DNA are extracted from an environmental sample, sequenced, analyzed and cloned in a laboratory-cultured host, *E. coli* ou *Komagataella phaffii P. pastoris*, and, after expression, the clones are subjected to screening for the desired property, in this case, the production of lipolytic enzymes and its biochemical and biophysical characterization (Piel, 2011).

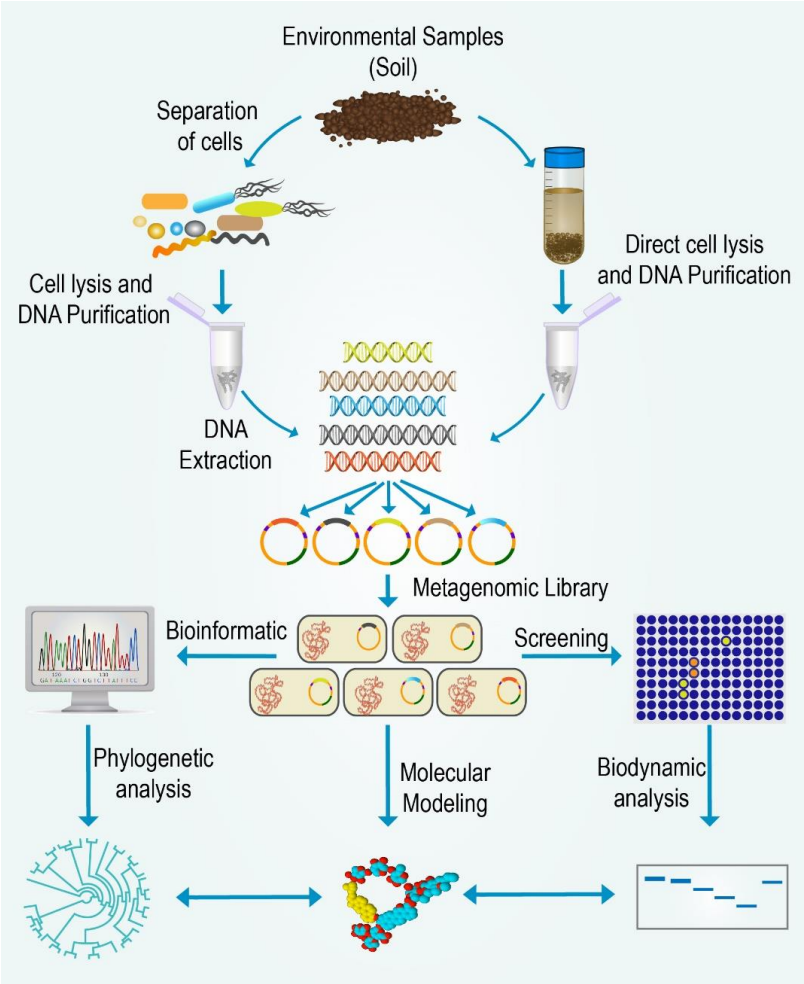


Figure 5 – Schematic representation of the metagenomic approach with the research strategy to access new biocatalysts from Cerrado soil.

Among the systems of heterologous expression, the prokaryotic and eukaryotic systems are observed. The prokaryotic hosts include bacteria and the eukaryotic hosts include yeasts, filamentous fungi, insect cells and mammal cells. In comparison with the insect cells and mammal cells, the bacteria, yeasts and filamentous fungi are usually more manageable in the laboratory and represent the heterologous platform most viable economically, thus more adequate for industrial applications.

The prokaryotic hosts are more used in the expression of the heterologous protein, since they have short growing timing and reach high cellular density at low cost. Besides, their genetics is well characterized, and a great number of cloning vectors and host mutant strips are available. It is worth highlighting, however, that they have inconveniences related to the inability to perform post translation modifications, including the correct folding of proteins, phosphorylation and glycosylation. Among the

gram-negative bacteria, the *E. coli* represent the host used with a range of molecular tools, including vectors commercially available for the construction of expression systems (Gopal and Kumar, 2013; Masuch et al., 2015).

The second most utilized prokaryotic organism to the expression of the recombinant protein is the gram-positive bacteria of the soil *B. subtilis*. In comparison to the *E. coli,* the *B. subtilis* has an advantage, as it secreted proteins at higher concentrations are directly in the environment (Borrelli and Trono, 2015).

The systems of eukaryotic expression, such as yeasts and filamentous fungi, has several advantages in terms of heterologous expression, showing a cellular organization that can accomplish post-translational modifications. The first yeast used to the heterologous expression of the recombinant protein was *S. cerevisiae.* This choice was based in the abundant data about its genetics and physiology, and its classified as safe (GRAS), allowing its wide use in the food industry (Valero, 2012). However, the protein expression in the *S. cerevisiae* faces several limitations, such as plasmid instability, low gains in production and low secretion capacity (Valero, 2012). For this reason, others yeasts were considered as alternative expression systems. Among such yeasts, the *Komagataella phaffii P. pastoris* has been the heterologous system most used for producing commercially relevant recombinant proteins, reforcing the fact that it uses methanol as the only carbon and energy source, in addition to the low secretion level from its own proteins.

This approach becomes the most suitable methodology to bypass the obstacle of cultivation, granting access to a number of genes that code new enzymes or more efficient enzymes from a biotechnology standpoint. In that sense, the biodiversity tools, the associated biochemical process and the heterologous expression systems are more and more studied considering industrial application. Our research group has already identified several genes with lipolytic activity, which grants them a huge potential in the industrial biotechnological process.

JUSTIFICATION

Soils allow a deep understanding of biodiversity involving microorganisms, loots of new molecules and especial lipolytic genes.

Several techniques can be used to study the material collected in the soils, despite the great challenges arising from this kind of study. One of the most efficient is the metagenomic technique, which allows accessing the microorganisms by independente cultivation approach, assisting in the identification of new genes and biocatalysts of interest.

Through this technique, it is possible to identify uncultured microorganisms, stimulating the study of the new lipolytic genes. As a consequence, it is possible to exploit the biotechnological potencial of this new enzymes, such as lipases, for example, in relevant industries.

General objective

- Structural and functional characterization of novel lipolytic enzymes from a Brazilian Cerrado soil metagenomic library

Specific objectives

- Cloning and sequencing of the gene and coding sequence (CDSs) of novel lipolytic enzymes from a Brazilian Cerrado soil metagenomic library;
- Heterologous expression of novel lipolytic enzymes in *E. coli;*
- Purification and functional characterization of recombinant lipolytic enzymes by one-dimensional electrophoresis and enzymatic essays;
- Biophysical characterization of recombinant lipolytic enzymes by circular dichroism and fluorescence spectroscopy;
- Modelling and structural studies of recombinant lipolytic enzymes.

Metagenomic libraries

The metagenomic library used in this work was previous constructed with Cerrado *stricto sensu* soil, as described (de Castro et al., 2011). Metagenomic DNA from the soil microbial community was partially digested with PstI in buffer H (Promega) and all DNA fragments between 2 and 8 kb were excised and extracted from the gel with GELase (Epicentre) according to the manufacturer's directions. Plasmid (pCF430) (Newman and Fuqua, 1999) vector was isolated and purified with QIAprep Miniprep (QIagen, Valencia, CA) and digested with 10 U PstI. The plasmid was dephosphorylated with shrimp alkaline phosphatase (Promega) according to the manufacturer's instructions. DNA ligation reactions between the soil microbial DNA inserts and the low copy number plasmid vector contained an insert/vector ratio of 3:1 and 3 U of T4 DNA ligase and buffer (Promega) was added to the reaction and incubated overnight at 16°C. Ligation products were transformed into electrocompetent *E. coli* TransforMax EPI300 and plated on LB medium containing tetracycline at 20 lg/ml. The libraries were stored at -80°C.

Phylogenetic analysis and screening of lipolytic enzymes

Functional screening of the metagenomic library was performed on Luria-Bertani (LB) agar (Jeon et al., 2009) containing 1% tributyrin, 40 μg/ml tetracycline, and 0.02% arabinose. Sequencing was performed using ABI PRISM 377 (Applied Biosystems), and coding sequence (CDS) prediction was performed using the PRODIGAL META procedure and default parameters (Hyatt et al., 2010). Translated sequences >150 amino acids were aligned against Swiss-Prot, TrEMBL, and UniRef100 (Bairoch and Boeckmann, 1994) using the DIAMOND sequence aligner with the BLASTp option (e-value <10), more-sensitive mode (Buchfink et al., 2014). To construct the phylogenetic tree, multiple alignments of the deduced protein sequences were performed in MAFFT v.7.388 (Katoh and Standley, 2013) using default parameters and manually edited. The phylogenetic tree was constructed in FastTree v.2.1 with BLOSUM62 matrix with 1,000 bootstrap (Price et al., 2010). Theoretical MW and isoelectric point (pI) were calculated using the Compute pI/Mw tool on the ExPASy server

(http://expasy.org/tools/pi_tool.html). Based on the results of functional screening and CDSs prediction, we chose to characterize LipX and lipY

From the functional screening and with the CDS prediction, we have continued with lipX gene and lipY to cloned into *E. coli* BL21 (DE3) stain and make the functional characterization.

Heterologous expression and purification of LipX and lipY

The *lipX and lipY* gene was PCR-amplified and the 966 and 924-bp amplicons were cloned into the pET24a expression vector (Novagen, USA), which carries an N-terminal His tag, to generate pET24a-lipX and pET24a-lipY (Apendix I).The vector was then transformed into electrocompetent *Escherichia coli* BL21 (DE3) cells, which were grown at 37°C for 20 h (200 rpm) in 500 ml LB medium containing 100 μg/ml Kanamicin. When OD_{600nm} reached 0.6–0.8, protein expression was induced with 0.1 mM isopropyl-β-D thiogalactopyranoside (IPTG) at 28°C. After twenty hours, the cells were centrifuged at 8,000 g for 10 min at 4°C and resuspended in 50 ml lysis buffer (10 Mm HEPES [pH 7.0], 10 mM imidazole, 0.03% [w/v] Triton X-100, 4 μg/ml lysozyme). After ultrasonication (40% amplitude, 10 cycles of 20 pulses with 10-s intervals) using an LB-750 Sonicator, the cells were centrifuged at 8,000 g for fourty-five minutes at 4°C. The protein of interest was purified using the MagneHis Protein Purification System (Promega, Madison, WI, USA) under native conditions with a modified protocol (dos Santos et al., 2015) and also proteins were purified by HiTrap HP Ni^2columns. The columns were placed in the ÄKTA™ and a blank run was made using binding buffer (20 mM Tris buffer, pH 7, containing 0.5 M NaCl and 20 mM imidazole), followed by elution buffer (20 mM Tris buffer, pH 7, containing 0.5 M NaCl and 0.5 M imidazole), and concluding with binding buffer. Five column volumes of buffer was used in each step. Proteins were eluted using a twenty volumes linear gradient going from 0 to 60% elution buffer, followed by a five volumes linear gradient up to 100% elution buffer. Columns were then washed with four volumes of 100% elution buffer and finally equilibrated with 5 CV of binding buffer Purity was confirmed by SDS-PAGE (13%).

Molar extinction coefficient and protein concentration

The molar extinction coefficient (ε) of LipX and lipY was determined by absorbance at 280 nm (A_{280}) as a function of protein concentration and previously described (Valle et al., 2018) An UV/Visible Spectrophotometer Jasco V-530 (Jasco Corporation, Tokyo, Japan) was used for (i) fixed point measurement and (ii) corrected absorbance considering light scattering at 350 nm (A_{350}) (Eq. 1).

$$Acorr_{280} = A_{280} - A_{350} \quad (1)$$

Protein concentration was determined by the Lowry method using a standard bovine serum albumin solution (0.1–1 mg/ml; Sigma, Steinheim, Germany) and measuring A_{620} after thirty minutes in a Multiskan EX microplate reader (Labsystems, Finland). The 200-µL reaction volume contained 4–300 nM LipX and LipY protein, 1 mM ρ-nitrophenyl (ρNP)-butyrate in isopropanol, and 50 mM sodium phosphate (pH 7.5) with 0.3% (v/v) Triton X-100 buffer (PBTx) at 30 °C.

Circular Dichroism (CD) measurements

The secondary structure profile of lipX and lipY at 2 mM sodium acetate, pH 5.0 and 2 mM Tris HCl pH 7.0 and pH 8.5, respectively, was analyzed by Circular Dichroism (CD) using a Jasco J-815 spectropolarimeter (Jasco Corporation, Tokyo, Japan) equipped with a Peltier-type temperature controller (Jasco Analytical Instruments, Japan). The Far-Ultraviolet (UV) CD spectra was recorded using a 0.05cm path length quartz cuvette with intervals of 0.2 nm at 25°C. Three consecutive measurements were registered, and the average spectrum was corrected for the baseline contribution of the buffers. The observed ellipticities were converted in a molar ellipticity [θ] (degree.cm^2.dmol^{-1}) based on molecular mass per 115 Da residue (Adler et al., 1973). The secondary structure of the enzyme was estimated using the CD spectra deconvolution (CDNN) version 2.1 software (Böhm et al., 1992). The thermal denaturation essays of the LipX and of lipY were performed recording the dichroic signal at 222 nm ([θ]222) by increasing the temperature from 25 to 95°C in the pHs 5.0, 7.0 and 8.5. The signals were collected at each 0.5ºC/min, interspersed by spectra obtained in the Far-UV region with intervals of 10ºC. The curves of thermal denaturation were defined considering the values of molar ellipticity ([θ]222nm) versus temperature. The melting temperature (Tm), where the protein unfolding occurs, was

calculated from the nonlinear fitting of unfolding curves using Origin software 8.0 program (Microcal Software Inc., Northampton, MA).

Fluorescence Spectroscopy Essays

The conformational changes of the lipX and lipY were analyzed by fluorescence spectroscopy using the Jasco FP-6500 Spectrofluorimeter (Jasco Analytical Instruments, Tokyo, Japan) connected to a Peltier type system controller (Jasco Analytical Instruments, Japan) with water circulation. The experiments were performed with the enzyme (0.061 mg/mL) in the different pHs, using 20 mM of sodium acetate (pH 4.5-5.5), 20 mM of Tris HCl (pH 6.0-9.0) and 20 mM of Glycine (pH 9.5). The fluorescence emission of tryptophan residues was obtained in the range of 300-400 nm, after the excitation at 295 nm at 25°C. Both the excitation and emission slits were fitted at 5 nm, respectively.

Enzyme activity essay

LipX and lipY enzymatic activity was measured by spectrophotometric detection against p-nitrophenyl (pNP) esters. The reaction mixture contained 1 mM pNP-butyrate in isopropanol and 50 mM PBS buffer (pH 7.0) with 0.3 % (v/v) Triton X-100. The enzyme activity was determined by kinetic monitoring of the product, p-nitrophenol, at 348 nm, the pH-independent isosbestic point of p-nitrophenol (Hriscu et al., 2013), with a SpectraMax M2e spectrophotometer (Molecular Devices, USA). All reactions were performed in triplicate. The effect of pH on enzyme activity was determined at 30 °C with 50 mM and each of the following buffer under standard essay conditions used the following overlapping buffer systems: KCl/HCl (pH 1–2.2), glycine/HCl (pH 2.2–3.6), sodium citrate (pH 3–6), sodium phosphate (pH 6–8), Tris/HCl (pH 7.6–9), CHES (pH 8.8–10), sodium carbonate (pH 9.6–10.6) and CAPS (pH 9.5 to 11). The optimal temperature was determined in a range of 20 to 65 °C (Dukunde et al., 2017). All reactions were performed in triplicate, and control reactions were performed without the enzymes for every measurement under the different conditions to subtract the values for the no enzymatic hydrolysis of the substrates. To determine the initial reaction velocities, the linear regressions and the mean standard deviations were calculated. All data obtained were analyzed using the Origin software v8.0 (OriginLab Corporation, EUA) and

whenever it was necessary ANOVA and Tukey's test at 5% probability were used to compare the treatment methods.

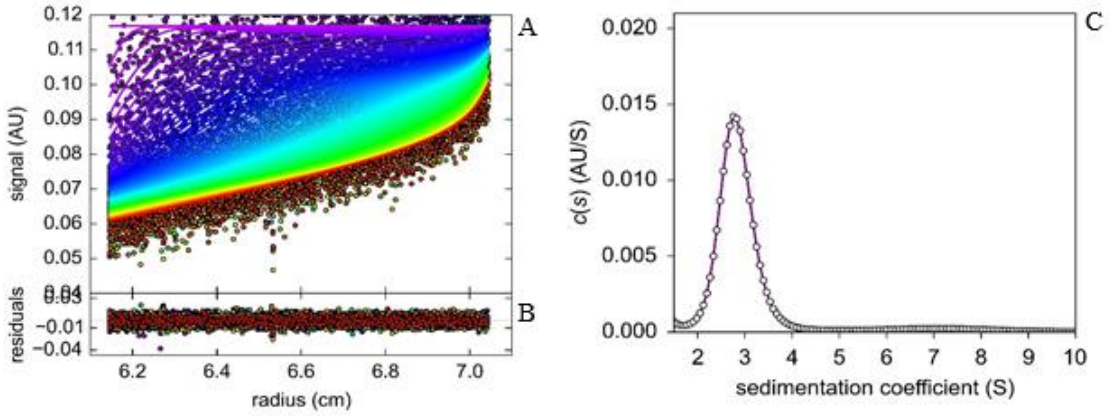
Modelling and structural studies of lipX and lipY

BLAST algorithm against protein data bank (PDB) was used to carry out the sequence homology searches (Altschul et al., 1997). The sequence and 3D structure of template proteins were extracted from the PDB database (Berman, 2000). Multiple sequence alignments of the target and template sequences were carried out using ClustalW and T-Coffee servers with default parameters (Larkin et al., 2007; Notredame et al., 2000). Based on high-resolution crystal structures of homologous proteins, 3D models of the lipX and lipY were built using the homology modeling software MODELLER 9v18 (Webb and Sali, 2014). Further, these models were improved by 3Drefiner (Bhattacharya and Cheng, 2013). The structures depiction were generated using PyMol software (DeLano, 2008). Stereochemical quality of structures was evaluated using Ramachandran plot obtained from RAMPAGE (Lovell et al., 2003). The preparation of the multiple sequence alignment figure and secondary structure depiction were carried out using ESPript version 3.0 (Robert and Gouet, 2014).

CHAPTER I

Online Resource 1 Annotation and taxonomic classification of the predicted coding regions of Clones W. The translated sequences of predicted CDSs (>150 aa) were aligned with protein sequences in the SwissProt, TrEMBL, and UniRef100 databases. Only the top hits of each alignment are shown.

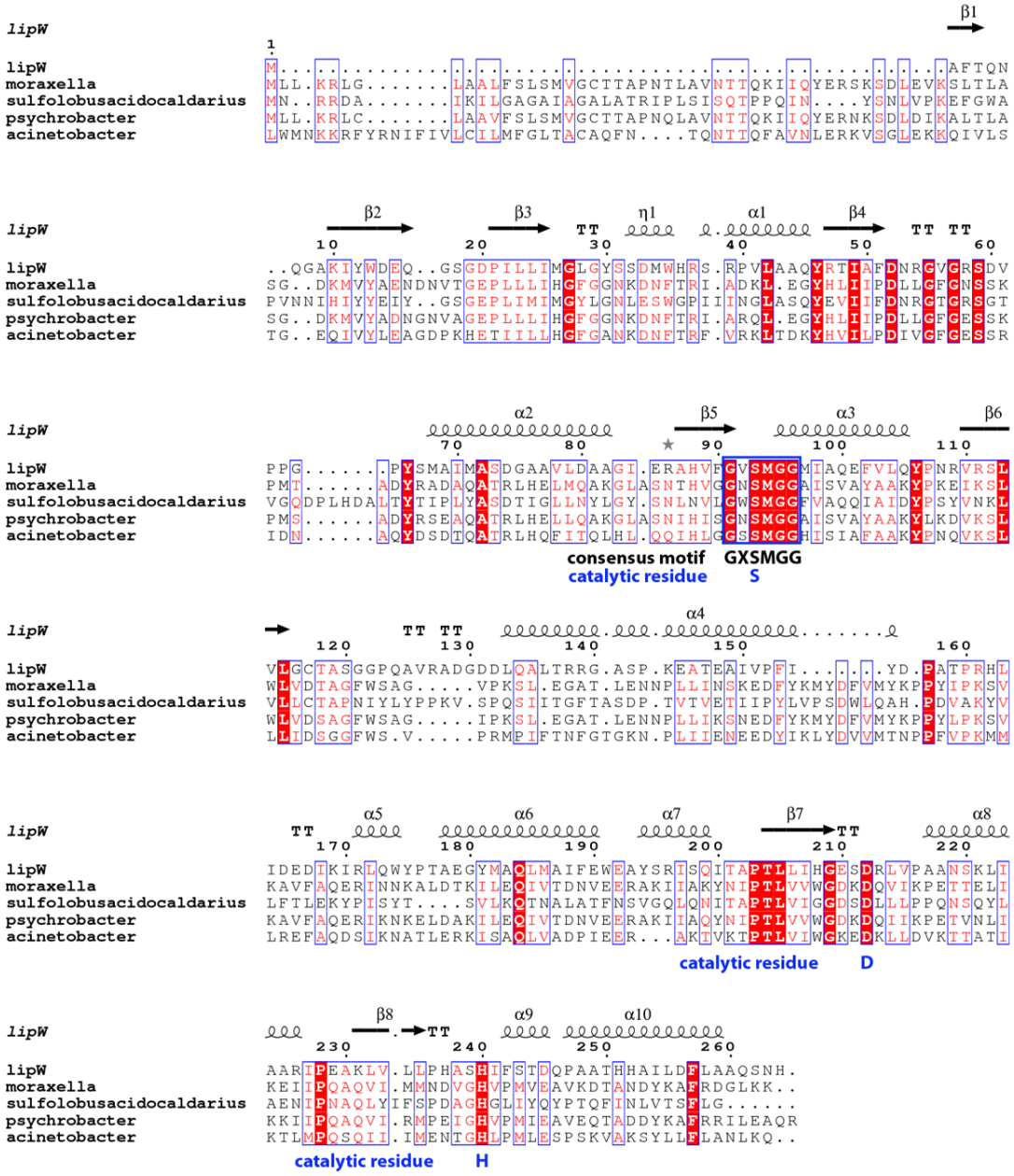
Uniprot - SwissProt								
CDS	**Start**	**End**	**Strand**	**E-value**	**bit score**	**ID**	**Annotation**	**Taxonomy**
CDS_W_5	2341	2970	1	4.5e-16	86.3	sp|A8IAD8|RUTD_AZOC5	Putative aminoacrylate hydrolase RutD	*Azorhizobium caulinodans*
CDS_W_8	3601	4188	-1	3.9e-06	53.1	sp|Q2RNG2|GLND_RHORT	Bifunctional uridylyltransferase/uridylyl-removing enzyme	*Rhodospirillum rubrum*
CDS_W_2	358	888	-1	2.8e-27	123.2	sp|Q9P5M9|MUG14_SCHPO	Meiotically up-regulated gene 14 protein	*Schizosaccharomyces pombe*
CDS_W_3	917	1378	-1	9.8e-13	74.7	sp|B3DMA2|ACD11_RAT	Acyl-CoA dehydrogenase family member 11	*Rattus norvegicus*
Uniprot - Trembl								
CDS_W_5	2341	2970	1	1.1e-38	168.7	tr|B0L3I5|B0L3I5_9BACT	Lipase/esterase	uncultured bacterium
CDS_W_8	3601	4188	-1	5.3e-40	172.9	tr|A0A1Q6X7Q5|A0A1Q6X7Q5_9BACT	Bifunctional uridylyltransferase/uridylyl-removing enzyme	Acidobacteria bacterium 13_2_20CM_57_7
CDS_W_2	358	888	-1	5.3e-55	222.6	tr|A0A1Q8BLQ8|A0A1Q8BLQ8_9CYAN	Class II aldolase/adducin family protein	Cyanobacteria bacterium 13_1_20CM_4_61_6
CDS_W_3	917	1378	-1	3.6e-15	90.1	tr|M7CMS2|M7CMS2_9ALTE	Acyl-CoA dehydrogenase family protein	Marinobacter santoriniensis NKSG1
UniRef100								
CDS_W_5	2341	2970	1	1.6e-38	168.7	UniRef100_B0L3I5	Lipase/esterase	uncultured bacterium
CDS_W_8	3601	4188	-1	7.7e-40	172.9	UniRef100_A0A1Q6X7Q5e	Bifunctional uridylyltransferase/uridylyl-removing enzym	Acidobacteria bacterium 13_2_20CM_57_7
CDS_W_2	358	888	-1	3.5e-55	223.8	UniRef100_UPI00047915CA	class II aldolase/adducin family protein	*Fischerella* sp. PCC 9605
CDS_W_3	917	1378	-1	5.2e-15	90.1	UniRef100_M7CMS2	Acyl-CoA dehydrogenase family protein	*Marinobacter santoriniensis* NKSG1



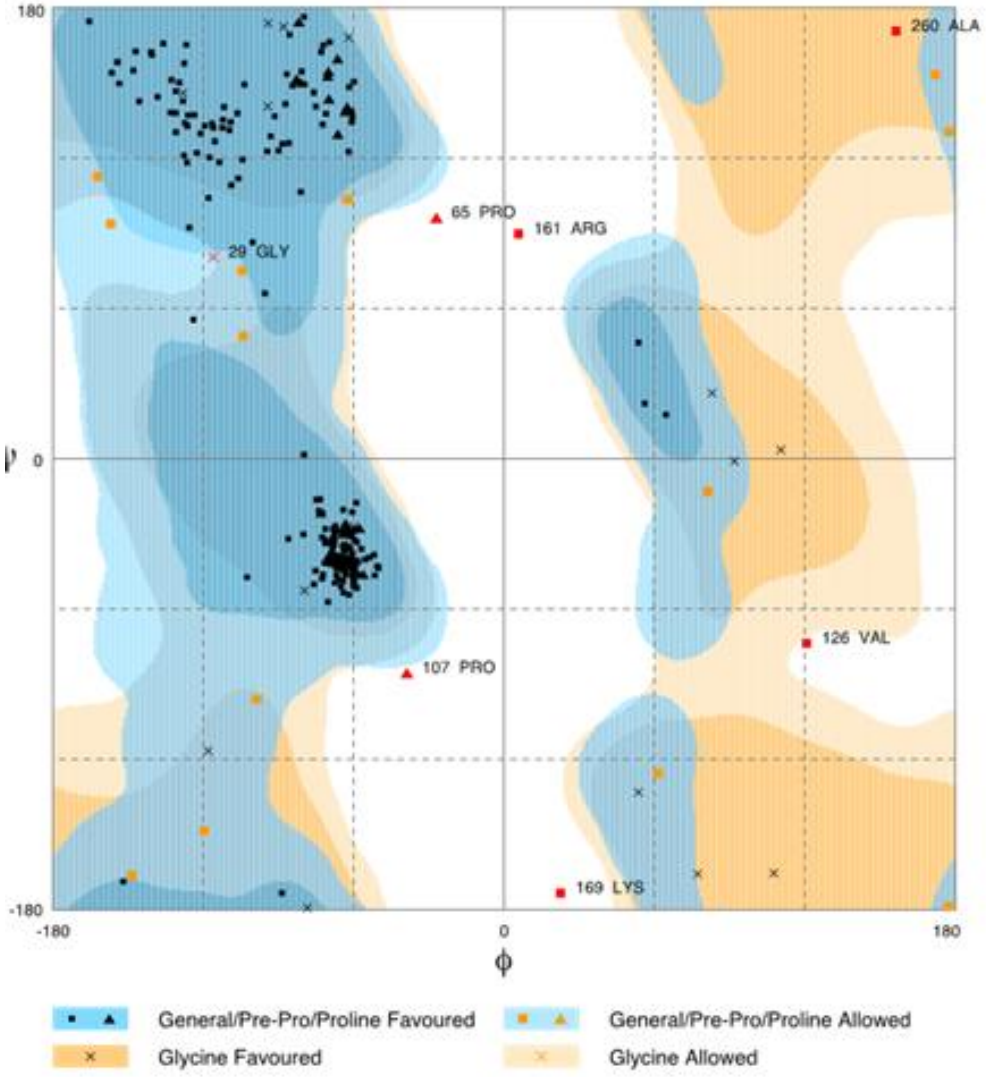
Online Resource 2 Sedimentation velocity. **a** Raw data of absorbance at 280 nm versus cell radius for 2.91 μM LipW. **b** Residuals plot supplied by SEDFIT showing goodness of fit. **(c)** Continuous sedimentation coefficient distribution, c(S) curve, obtained with a regularization procedure from data shown in panel **a** with a confidence level of 0.95 using SEDFIT and frictional ratios of 1.20–1.31. The partial specific volume (υ) 0.735614 ml/g for LipW was determined using SEDNTERP. Solvent (water) density (ρ= 0.99823 g/ml) and viscosity (η=0.01002 poise) were also determined by SEDNTERP. Circles represent experimental data, and the solid line represents the best fit to the Lamm equation supplied by SEDFIT. Similar results were obtained for 2.14 and 1.56 μM LipW.

Online Resource 3 Molecular and biophysical parameters of LipW.

Parameters	**LipW 2.91 μM**	**LipW 2.14 μM**	**LipW 1.56 μM**
Theoretical MW***** (Da)		29,530.55	
MW (Da)	29,338.67 ± 113.37	29,411.98 ± 152.27	29,380.16 ± 184.74
$S_{20,w}$ (*S*)	3.10 ± 0.01	3.60 ± 0.01	2.84 ± 0.01
Stokes radius (nm)	2.58	2.58	2.83



Online Resource 4 Amino acid sequence alignment of LipW with family V lipases. The G-X-S-M-G-G consensus motif of family V is shown, and the catalytic triad residues are denoted with asterisks.



Number of residues in favoured region (~98.0% expected)	: 242 (92.4%)
Number of residues in allowed region (~2.0% expected)	: 13 (5.0%)
Number of residues in outlier region	: 7 (2.7%)

RAMPAGE by Paul de Bakker and Simon Lovell available at http://www-cryst.bioc.cam.ac.uk/rampage/

Please cite: S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson & D.C. Richardson (2002) Structure validation by Cα geometry: φ/ψ and Cβ deviation. *Proteins: Structure, Function & Genetics*, **50**:437-450

Online Resource 5 Ramachandran plot of the LipW model.

CHAPTER II

Novel lipolytic enzymes from a Brazilian Cerrado soil metagenomic library cloned into *E. coli* BL21 (DE3) (The results of this chaper will be submitted to PLOsOne).

Results and Discussion

Phylogenetic Analysis

The initial prediction of CDS located in the metagenomic Clones X and Y was accomplished using Prodigal meta procedure and default parameters (Appendices 1). Clone X is a 13.650 bp fragment and after analysis of their sequence, eight CDSs were identified. The alignment against SwissProt, Trembl and UniRef100 reference databases identified CDS_X_14 as a Glutamine synthetase from *Bradyrhizobium sp* in all tree databases. SwissProt best match of CDS_X_10 was a carboxylesterase from mycobacterium tuberculosis, with Trembl as an acetylhydrolase from *Bradyrhizobium sp* and with UniRef100 as an alpha/beta hydrolase from *Bradyrhizobium sp.* CDS_X_13 with SwissProt was found as an *Exopolysaccharide* production negative regulator from *Rhizobium meliloti,* with Trembl as a Secretion pathway protein E from *Bradyrhizobium sp* and with UniRef100 as a General secretion pathway protein E from *Bradyrhizobium sp* as well. CDS_X_7 with SwissProt was found as an *Exopolysaccharide* production negative regulator from *Aspergillus oryzae* with Trembl as an uncharacterized protein from *Bradyrhizobium sp* and with UniRef100 as an uncharacterized protein from *Bradyrhizobium sp*. as well. CDS_X_6 alignment only against Trembl as an uncharacterized protein from *Bradyrhizobium sp* and with UniRef100 as a domain-containing protein from *Bradyrhizobium sp*. CDS_X_4 with SwissProt was found as an NAD(P)H-dependent FMN reductase from S*chizosaccharomyces* pombe with Trembl as an NAD(P)H-dependent FMN reductase from *Bradyrhizobium sp* and with UniRef100 as an NADPH-dependent oxidoreductase from *Bradyrhizobium sp* as well. CDS_X_2 with SwissProt was found as an uncharacterized protein from *Sinorhizobium fredii* with Trembl as a transposase from *Acidithiobacillus thiooxidans* and with UniRef100 as a Transposase from *Acidithiobacillus thiooxidans* as well. CDS_X_1 alignment only against Trembl as an Uncharacterized protein from Polaromonas sp as a Hypothetical protein from *Methylocapsa palsarum.*

Clone Y is 6.650 base pairs and, after the analysis of their sequence, eight CDSs were identified. The alignment against SwissProt, Trembl and UniRef100 reference databases found CDS_Y_4 with SwissProt. CDS_Y_4 was also found, using SwissProt, as a Monooxygenase from *Pseudomonas aeruginosa,* and with Trembl as a

Monooxygenase from *Acidobacteria bacterium*. Using UniRef100, CDS_Y_4 was found as a Monooxygenase from *Acidobacteria bacterium* as well. CDS_Y_8 with SwissProt was found as an Aminomethyltransferase from *Geobacillus sp*., with Trembl as an Uncharacterized protein from *Acidobacteria bacterium* and, with UniRef100 as an Uncharacterized protein from Unclassified *Acidobacteria*. CDS_Y_2 with SwissProt was found as an Acetyl-hydrolase from Streptomyces hygroscopicus, with Trembl as an Uncharacterized protein from *Acidobacteria* bacterium and with UniRef100 as an Uncharacterized protein from Acidobacteria bacterium as well. CDS_Y_6 with SwissProt was found as a Beta-glucanase from *Rhodothermus marinus*, with Trembl as an Uncharacterized protein from *Acidobacteria bacterium* and with UniRef100 as an Uncharacterized protein from *Acidobacteria bacterium* as well. CDS_Y_1 with SwissProt was found as a Dehydrogenase from *Bacillus subtilis*, with Trembl as a Dehydrogenase from *Acidobacteria bacterium* and with UniRef100 as an Uncharacterized protein from *Acidobacteria bacterium*. CDS_Y_5 with SwissProt was found as an Oxidoreductase from *Koribacter versatilis*, with Trembl as an Oxidoreductase from *Acidobacteria bacterium* and with UniRef100 as an Uncharacterized protein from *Acidobacteria bacterium*. CDS_Y_7 with SwissProt was found as a Protein SPy from *Streptococcus pyogenes*, with Trembl as an Uncharacterized protein from *Acidobacteria bacterium* and with UniRef100 as an Uncharacterized protein from *Acidobacteria bacterium* as well. CDS_Y_3 with SwissProt was found as a Monooxygenase from Pseudomonas aeruginosa, with Trembl as a Monooxygenase from *Acidobacteria bacterium* and with UniRef100 as a Monooxygenase *from Acidobacteria bacterium* as well.

Within the identified CDSs on the metagenomics cloneX and cloneY we select clones with lipolytic activity for further analysis, following some criteria: the alignment against Trembl and UniRef100 reference, if they share the typical catalysis pentapeptide GXSXG in protein sequence and have a common catalytic triad.

The lines highlighted in red show CDS with lipolytic activity, which were named lipX, lipY and lipW and sub-cloned into a pET24a and pET21a respectively vectors, with subsequent transformation in *Escherichia coli* (DE3) cells.

Table 1 *– Annotation and taxonomic classification of the predicted coding regions in Clone X, Y and W sequence. The translated sequences of predicted CDSs (> 150 aa) were aligned with protein sequences of SwissProt database.*

					Uniprot - SwissProt			
CDS	Start	End	Strand	E-value	bit score	ID	Annotation	Taxonomy
CDS_X_14	11454	12461	-1	3.1e-176	619.0	sp|P04772|GLNA2_BRADU	Glutamine synthetase	Bradyrhizobium diazoefficiens
CDS_X_10	8399	9361	1	6.6e-35	149.4	sp|P9WK86|NLHH_MYCTO	Carboxylesterase NlhH	Mycobacterium tuberculosis
CDS_X_13	10482	11402	-1	7.5e-04	46.2	sp|Q52926|EXOR_RHIME	Exopolysaccharide production negative regulator	Rhizobium meliloti
CDS_X_7	6505	7350	-1	8.9e-28	125.6	sp|Q2UPB6|ACLN_ASPOR	Aspirochlorine biosynthesis protein N	Aspergillus oryzae
CDS_X_4	3578	4168	-1	9.3e-08	58.5	sp|Q9USJ6|FMNR_SCHPO	NAD(P)H-dependent FMN reductase C4B3.06c	Schizosaccharomyces pombe
CDS_X_2	1658	2137	1	7.1e-14	78.6	sp|P55614|Y4PE_SINFN	Uncharacterized protein y4pE/y4sA	Sinorhizobium fredii
CDS_Y_4	1965	3020	1	4.6e-90	332.8	sp|Q9I3H5|BVMO_PSEAE	Baeyer-Villiger monooxygenase	Pseudomonas aeruginosa
CDS_Y_8	5335	6339	-1	1.9e-24	114.8	sp|C5D4A2|GCST_GEOSW	Aminomethyltransferase	Geobacillus sp.
CDS_Y_2	689	1609	1	1.7e-48	194.5	sp|Q01109|BAH_STRHY	Acetyl-hydrolase	Streptomyces hygroscopicus
CDS_Y_6	3765	4652	1	2.3e-50	200.7	sp|P45798|GUB_RHOMR	Beta-glucanase	Rhodothermus marinus
CDS_Y_1	2	643	1	1.4e-25	117.9	sp|O35017|YOGA_BACSU	Uncharacterized zinc-type alcohol dehydrogenase-like protein YogA	Bacillus subtilis
CDS_Y_5	3203	3742	1	6.4e-72	271.6	sp|Q1IPE7|NUOB1_KORVE	NADH-quinone oxidoreductase subunit B 1	Koribacter versatilis
CDS_Y_7	4767	5255	1	1.2e+00	34.7	sp|Q9A0E3|Y811_STRP1	UPF0342 protein SPy_0811/M5005_Spy0626	Streptococcus pyogenes
CDS_Y_3	1609	1944	1	9.6e-26	117.5	sp|Q9I3H5|BVMO_PSEAE	Baeyer-Villiger monooxygenase	Pseudomonas aeruginosa
CDS_W_5	2341	2970	1	4.5e-16	86.3	sp|A8IAD8|RUTD_AZOC5	Putative aminoacrylate hydrolase RutD	Azorhizobium caulinodans
CDS_W_8	3601	4188	-1	3.9e-06	53.1	sp|Q2RNG2|GLND_RHORT	Bifunctional uridylyltransferase/uridylyl-removing enzyme	Rhodospirillum rubrum
CDS_W_2	358	888	-1	2.8e-27	123.2	sp|Q9P5M9|MUG14_SCHPO	Meiotically up-regulated gene 14 protein	Schizosaccharomyces pombe
CDS_W_3	917	1378	-1	9.8e-13	74.7	sp|B3DMA2|ACD11_RAT	Acyl-CoA dehydrogenase family member 11	Rattus norvegicus

Table 2 *– Annotation and taxonomic classification of the predicted coding regions in Clone X, Y and W sequence, aligned with protein sequences of Trembl database.*

					Uniprot - Trembl			
CDS	**Start**	**End**	**Strand**	**E-value**	**bit score**	**ID**	**Annotation**	**Taxonomy**
CDS_X_14	11454	12461	-1	6.9e-189	668.3	tr|H0SF60|H0SF60_BRAS3	Glutamine synthetase	Bradyrhizobium sp.
CDS_X_10	8399	9361	1	6.1e-142	512.3	tr|A0A1B9YTX8|A0A1B9YTX8_9BRAD	Acetylhydrolase	Bradyrhizobium sp.
CDS_X_13	10482	11402	-1	2.8e-80	307.4	tr|U1GVC1|U1GVC1_9BRAD	General secretion pathway protein E	Bradyrhizobium sp.
CDS_X_6	4602	5462	-1	4.8e-82	313.2	tr|A0A0R3LZJ8|A0A0R3LZJ8_9BRAD	Uncharacterized protein	Bradyrhizobium jicamae
CDS_X_7	6505	7350	-1	1.4e-134	487.6	tr|A0A1C3X4B4|A0A1C3X4B4_9BRAD	Uncharacterized protein	Bradyrhizobium sp.
CDS_X_4	3578	4168	-1	3.8e-94	352.8	tr|A0A1N6KKH9|A0A1N6KKH9_9BRAD	NAD(P)H-dependent FMN reductase	Bradyrhizobium erythrophlei
CDS_X_1	1116	1610	1	4.7e-21	109.8	tr|A0A1S8FNS3|A0A1S8FNS3_9BURK	Uncharacterized protein	Polaromonas sp.
CDS_X_2	1658	2137	1	1.2e-37	164.9	tr|A0A1C2JCA4|A0A1C2JCA4_ACITH	Transposase	Acidithiobacillus thiooxidans
CDS_Y_4	1965	3020	1	3.4e-162	579.7	tr|A0A1Q7K2H3|A0A1Q7K2H3_9BACT	4-hydroxyacetophenone monooxygenase	Acidobacteria bacterium
CDS_Y_8	5335	6339	-1	1.3e-134	488.0	tr|A0A1Q6X292|A0A1Q6X292_9BACT	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_2	689	1609	1	7.6e-126	458.8	tr|A0A1Q6YE25|A0A1Q6YE25_9BACT	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_6	3765	4652	1	6.6e-135	488.8	tr|A0A1Q7K2K1|A0A1Q7K2K1_9BACT	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_1	2	643	1	9.8e-96	358.2	tr|A0A1Q7K2I9|A0A1Q7K2I9_9BACT	Alcohol dehydrogenase	Acidobacteria bacterium
CDS_Y_5	3203	3742	1	2.8e-96	359.8	tr|A0A1Q7KYP2|A0A1Q7KYP2_9BACT	NADH-quinone oxidoreductase subunit B	Acidobacteria bacterium
CDS_Y_7	4767	5255	1	2.0e-69	270.4	tr|A0A1Q6YFR6|A0A1Q6YFR6_9BACT	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_3	1609	1944	1	2.9e-43	183.0	tr|A0A1Q7CH14|A0A1Q7CH14_9BACT	4-hydroxyacetophenone monooxygenase	Acidobacteria bacterium
CDS_W_5	2341	2970	1	1.1e-38	168.7	tr|B0L3I5|B0L3I5_9BACT	Lipase/esterase	uncultured bacterium
CDS_W_8	3601	4188	-1	5.3e-40	172.9	tr|A0A1Q6X7Q5|A0A1Q6X7Q5_9BACT	Bifunctional uridylyltransferase/uridylyl-removing enzyme	Acidobacteria bacterium 13_2_20CM_57_7
CDS_W_2	358	888	-1	5.3e-55	222.6	tr|A0A1Q8BLQ8|A0A1Q8BLQ8_9CYAN	Class II aldolase/adducin family protein	Cyanobacteria bacterium 13_1_20CM_4_61_6
CDS_W_3	917	1378	-1	3.6e-15	90.1	tr|M7CMS2|M7CMS2_9ALTE	Acyl-CoA dehydrogenase family protein	Marinobacter santoriniensis NKSG1

Table 3 *– Annotation and taxonomic classification of the predicted coding regions in Clone X, Y and W sequence, aligned with protein sequences of UniRef100 database.*

					UniRef100		
CDS	**Start**	**End**	**Strand**	**E-value**	**bit score**	**ID**	**Annotation**	**Taxonomy**
CDS_X_14	11454	12461	-1	7.0e-190	672.2	UniRef100_UPI000400DA42	glutamine synthetase	Bradyrhizobium sp. Tv2a-2
CDS_X_10	8399	9361	1	3.8e-153	550.1	UniRef100_UPI000464C94C	alpha/beta hydrolase	Bradyrhizobium sp. ARR65
CDS_X_13	10482	11402	-1	4.1e-80	307.4	UniRef100_U1GVC1	General secretion pathway protein E	Bradyrhizobium sp. DFCI-1
CDS_X_6	4602	5462	-1	2.0e-97	364.8	UniRef100_UPI00040FCF2C	DUF937 domain-containing protein	Bradyrhizobium sp. Tv2a-2
CDS_X_7	6505	7350	-1	2.1e-147	530.8	UniRef100_UPI000466992D6	hypothetical protein	Bradyrhizobium sp. ARR65
CDS_X_4	3578	4168	-1	2.5e-94	354.0	UniRef100_UPI000409A0E6	NADPH-dependent oxidoreductase	Bradyrhizobium sp. Ai1a-2
CDS_X_1	1116	1610	1	1.4e-26	128.6	UniRef100_UPI000B80D7C3	hypothetical protein	Methylocapsa palsarum
CDS_X_2	1658	2137	1	1.7e-37	164.9	UniRef100_A0A1C2JCA4	Transposase n=3	Acidithiobacillus thiooxidans
CDS_Y_4	1965	3020	1	4.9e-162	579.7	UniRef100_A0A1Q7K2H3	4-hydroxyacetophenone monooxygenase	Acidobacteria bacterium
CDS_Y_8	5335	6339	-1	1.9e-134	488.0	UniRef100_A0A1Q6X292	Uncharacterized protein	unclassified Acidobacteria
CDS_Y_2	689	1609	1	1.1e-125	458.8	UniRef100_A0A1Q6YE25	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_6	3765	4652	1	9.6e-135	488.8	UniRef100_A0A1Q7K2K1	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_1	2	643	1	1.4e-95	358.2	UniRef100_A0A1Q7K2I9	Alcohol dehydrogenase	Acidobacteria bacterium
CDS_Y_5	3203	3742	1	4.1e-96	359.8	UniRef100_A0A1Q7KYP2	NADH-quinone oxidoreductase subunit B	Acidobacteria bacterium
CDS_Y_7	4767	5255	1	3.0e-69	270.4	UniRef100_A0A1Q6YFR6	Uncharacterized protein	Acidobacteria bacterium
CDS_Y_3	1609	1944	1	4.3e-43	183.0	UniRef100_A0A1Q7CH14	4-hydroxyacetophenone monooxygenase	Acidobacteria bacterium
CDS_W_5	2341	2970	1	1.6e-38	168.7	UniRef100_B0L3I5	Lipase/esterase	uncultured bacterium
CDS_W_8	3601	4188	-1	7.7e-40	172.9	UniRef100_A0A1Q6X7Q5e	Bifunctional uridylyltransferase/uridylyl-removing enzym	Acidobacteria bacterium 13_2_20CM_57_7
CDS_W_2	358	888	-1	3.5e-55	223.8	UniRef100_UPI00047915CA	class II aldolase/adducin family protein	Fischerella sp. PCC 9605
CDS_W_3	917	1378	-1	5.2e-15	90.1	UniRef100_M7CMS2	Acyl-CoA dehydrogenase family protein	Marinobacter santoriniensis NKSG1

To analyze the phylogenetic relationship of lipX and lipY with representative members of bacterial lipolytic enzymes, a phylogenetic tree was built. LipX and lipY clustered with lipolytic enzymes from Family IV (Figure 6), being closely related to a lipase from *Pseudomonas sp.* (accession number AAC38151), *Cupriavidus necator* (accession number AAC 41424) and *Moraxella sp. (*accession number CAA37862*).* These enzymes show activities against p-nitrophenyl esters of fatty acids with short to medium chains (Choo et al., 1998).

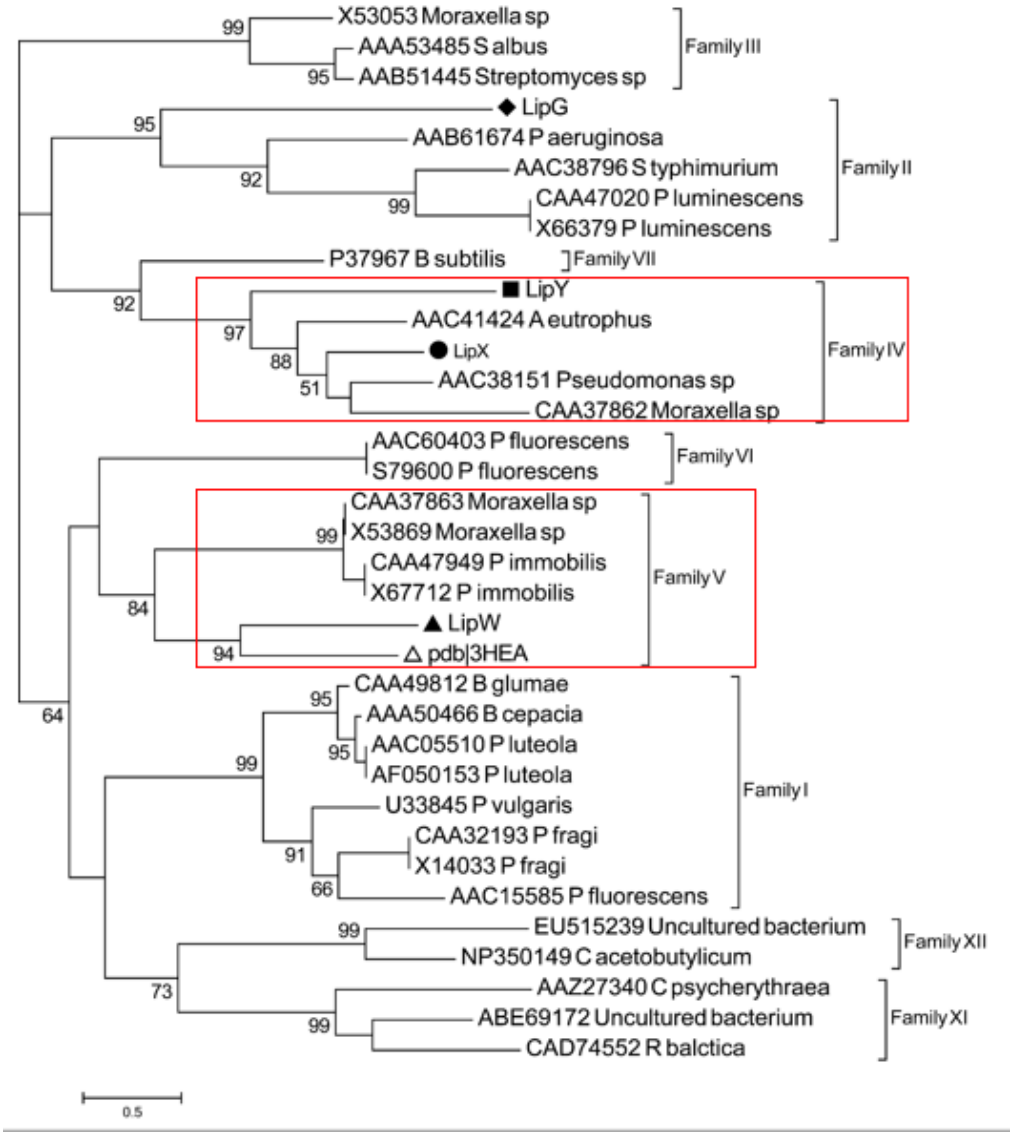


Figure 6 – Phylogenetic analysis of bacterial lipase families.
The tree was constructed by PhyML method using BLOSUM62 matrix. The analyses were bootstrapped (1000 replications), and only values greater than 50 % are shown. The scale bar indicates the number of amino acid substitutions per site. Family namer is indicated before the name of the species from which the enzyme originated

The classification of enzymes based on sequence provides an indication of the evolutionary relationship between the enzymes (Tyzack et al., 2017), it is a way of characterizing the sequence family and serves as recognition pattern. Because of that, characterizing metagenomic CDS from soil has some challenges. The most impactful is low sequence similarities in existing databases, which shows as hypothetical or putative proteins, conserved domains not known and only some solved three-dimensional structures. However, functional screening, structural and biochemical, and biophysical characterization provides important signals about these sequences, with the substrate used in the initial essay being a strong indication for the classification of the CDS inside the insert.

The predicted molecular weight and isoelectric point for the mature lipX were estimated as 35 kDa and 5.3, respectively. (Table 4).The number of negatively charged residues (Asp + Glu) were estimated as forty and the number of positively charged residues (Arg + Lys) as thirty-one. The lipX formula was estimated as C1551H2421N447O460S6 with a total number of atoms of 4885. The predicted molecular weight and isoelectric point for the mature lipY were estimated as 32 kDa and 7.8, respectively, the number of negatively charged residues (Asp + Glu) were estimated as thirty-three and the number of positively charged residues (Arg + Lys) as thirty-four. The lipY formula was estimated as C1537H2397N431O426S8 with a total number of atoms of 4799.

Table 4 *– BLASTP annotation of metagenomic clones CDS with lipolytic activity*

Subclone	**AA**	**BestHits**	**(kDa)/iP**	**% Similarity**	**Preserved Domain**	**Access N**
lipX	320	*[Bradyrhizobium* sp].	*35/5,3*	79%	α,βAbhydrolase	*WP_024510238*
lipY	306	[*Acidobacteria bacterium*]	*33/7,8*	75%	α,βAbhydrolase	OLB36079

Regarding enzyme identified as lipX, the aligned with a metagenomic acetylhydrolase from *Bradyrhizobium sp* family (accession number WP_024510238) was its best hit, sharing 79% of sequence similarity (Figure 7).

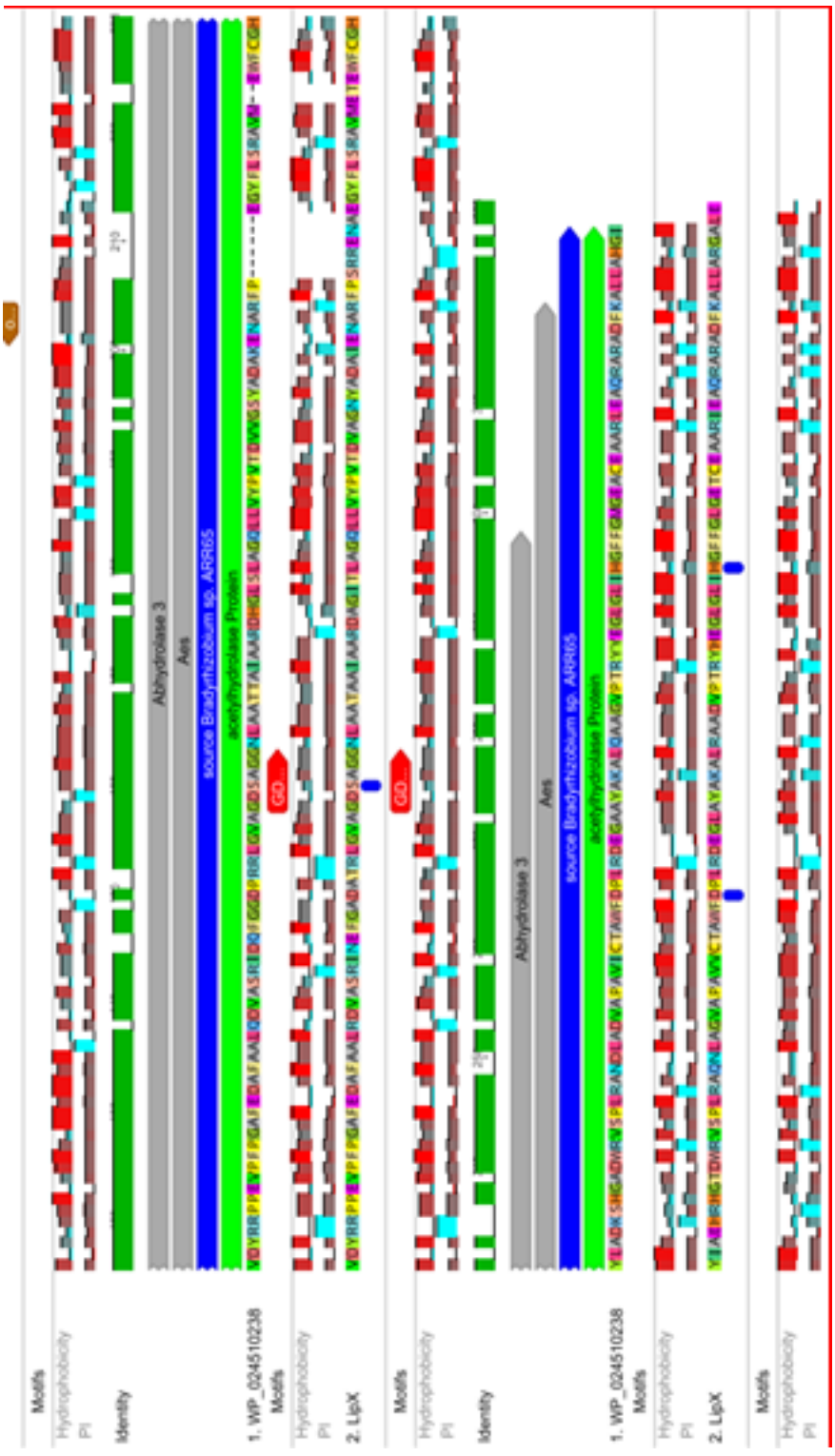


Figure 7 – Analysis of a LipX subclone with besthit WP_024510238:
the alignment, the typical catalysis pentapeptide: GxSxGG- G154, G158, G159) (red arrows), common catalytic triad S156, D259 and H289 (blue spots) and oxyanion hole: HGG - G83, G84 (brown arrow).

Sequence alignment of lipX with the best hit revealed the conserved pentapeptide motif well conserved Gly–X–Ser–X–Gly found in many esterases and the common catalytic triad S156, D259 and H289 (blue spots). Additionally, the evidence of the highly conserved HGGG: G83, G84 (brown arrow) motif in the LipX sequence upstream of the serine motif, this motif is involved in hydrogen-bonding interactions that stabilize the oxyanion hole and plays a role in catalysis (Dukunde et al., 2017).

Regarding enzyme identified as lipY, alignment with a metagenomic hypothetic protein from *Acidobacteria bacterium* (accession number OLB36079) was its best hit, sharing 75% of sequence similarity (Figure 8). Sequence alignment of lipY revealed a conserved pentapeptide motif well conserved Gly–X–Ser–X–Gly found in many esterases and the common catalytic triad S156, D259 and H289 (blue spots) and oxyanion hole: HGG - G83, G84 (brown arrow).

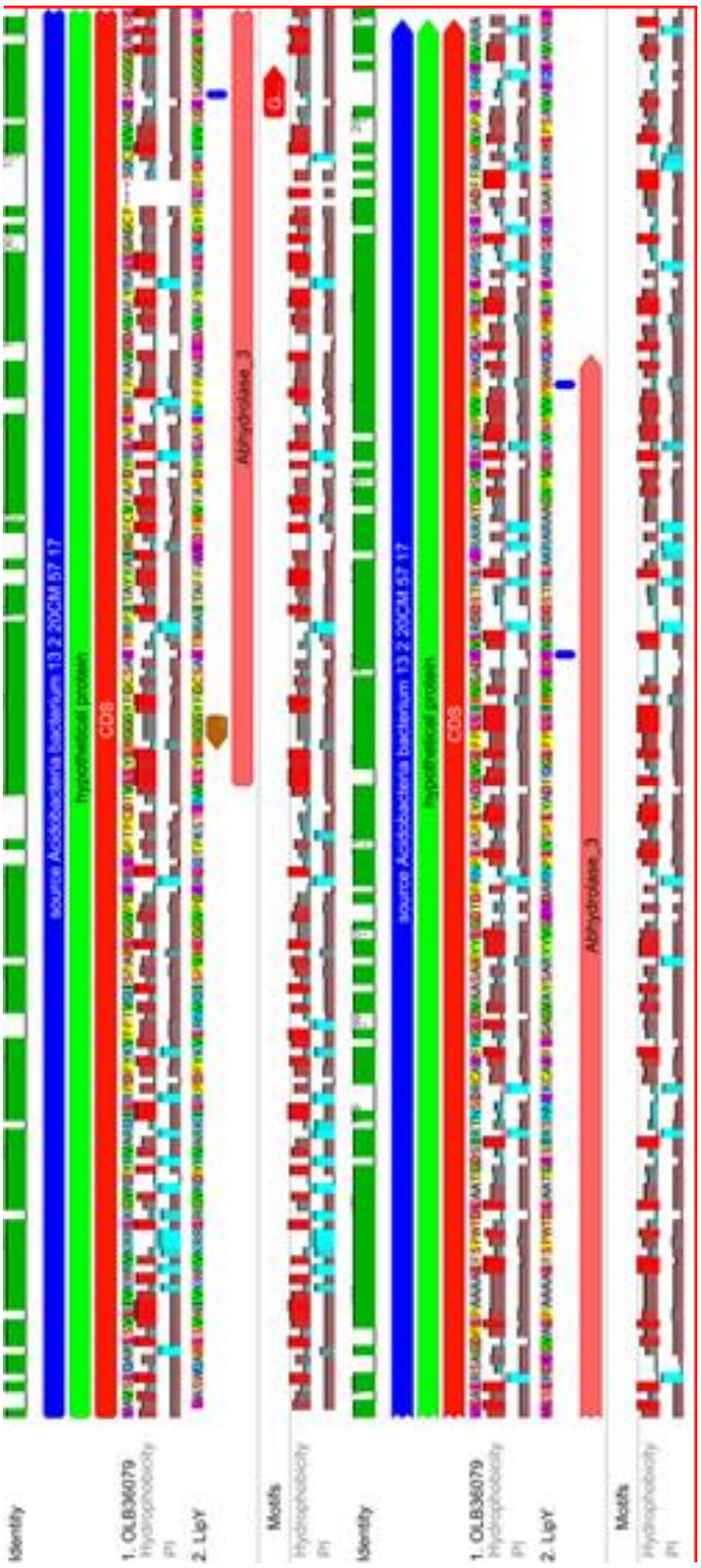


Figure 8 – Analysis of a LipY subclone with besthit OLB36079:
the alignment, the typical catalysis pentapeptide: GxSxGG- G154, G158, G159) (red arrows), common catalytic triad S156, D259 and H289 (blue spots) and oxyanion hole: HGG - G83, G84 (brown arrow).

Expression in E.coli recombinant lipX and lipY protein and purification

The recombinant lipX and lipY proteins were fused to a N-terminal His tag that was obtained from BL21 (DE3) *E.coli* strain and transformed with pET24a-lipX and pET24a-lipY expression vectors. They showed activity after the transformation into tributyrin plate (Figure 9).

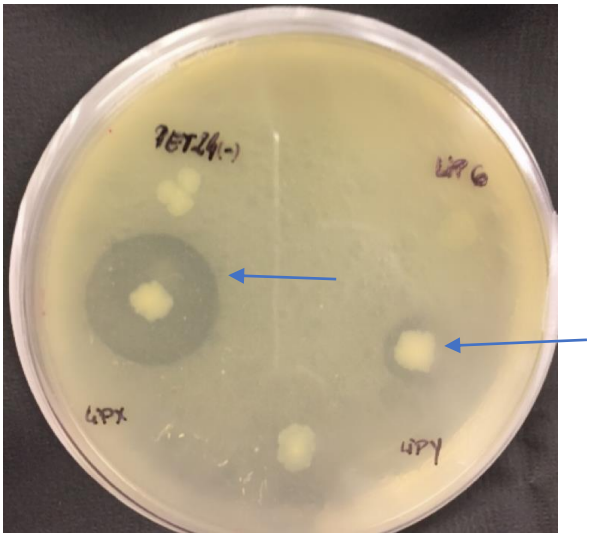


Figure 9 – The recombinant lipX and lipY protein expressed in the BL21 (DE3) E. coli strain on pET24a expression vectors after the plating into tributyrin containing mediume showing lypolitic activity.

The purification of lipX revealed a 35-kDa recombinant protein, as predicted by the sequence data (Figure 10).

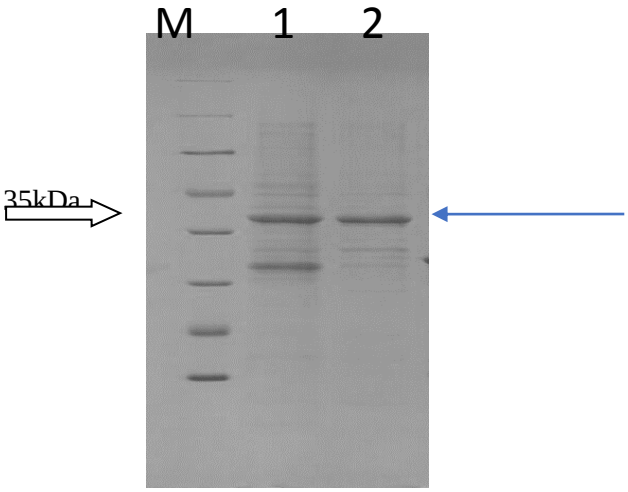


Figure 10 – Purification of recombinant lipX. SDS-PAGE (13 %) M: molecular weight standard *(Thermo Scientific Lane1Ni-NTA affinity chromatography: purification fraction of lipX from E. coli BL21(DE3) cells carrying the pET24a-lipX vector using Akta. Lane2 Ni-NTA*

The purification of lipY revealed a 32-kDa recombinant protein, as predicted by the sequence data (Figure 11)

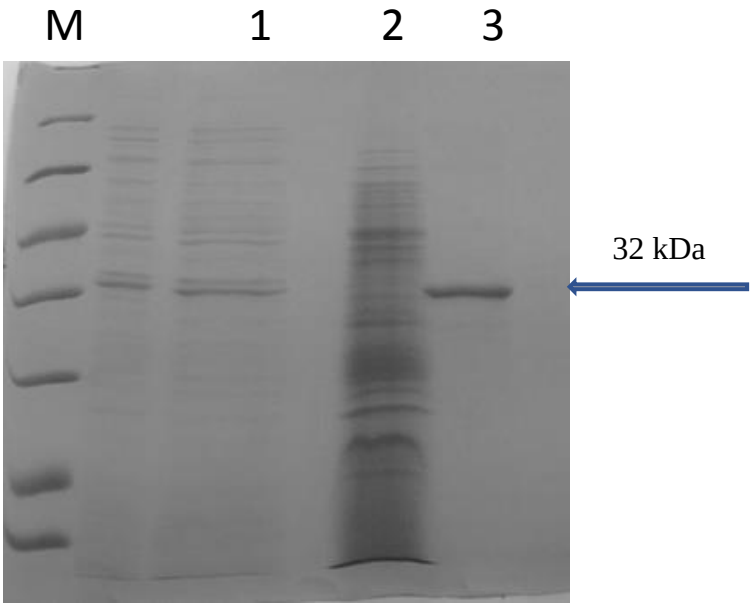


Figure 11 – Expression, purification of recombinant lipY SDS-PAGE (13 %) of the expression and Ni-NTA affinity chromatography purification fractions of lipY from E. coli BL21(DE3) cells carrying the pET24a-lipYvector*. M: molecular weight standard (Thermo Scientific Lane 1: LipY induction with 0.1mM IPTG. Lane 2: crude extract of the induced cells. Lane 3: purified LipY*

LipX and lipY enzyme activity

Enzyme activity of lipX was observed from pH 3.0–10.0, with optimal activity at pH 8.5 (Fig. 12A). Thus, lipX is not only active over a wide range of pH values but it is active under neutral as well as alkaline conditions. The results showed that lipX is more efficient in basic pH (8.5 and 9.5) relative neutral pH. Within the temperature range of 25°C–95°C, maximum activity was observed at 55°C; relative enzyme activity was >60% from 30°C–45°C and approximately 60% at 60°C (Fig. 12 B).

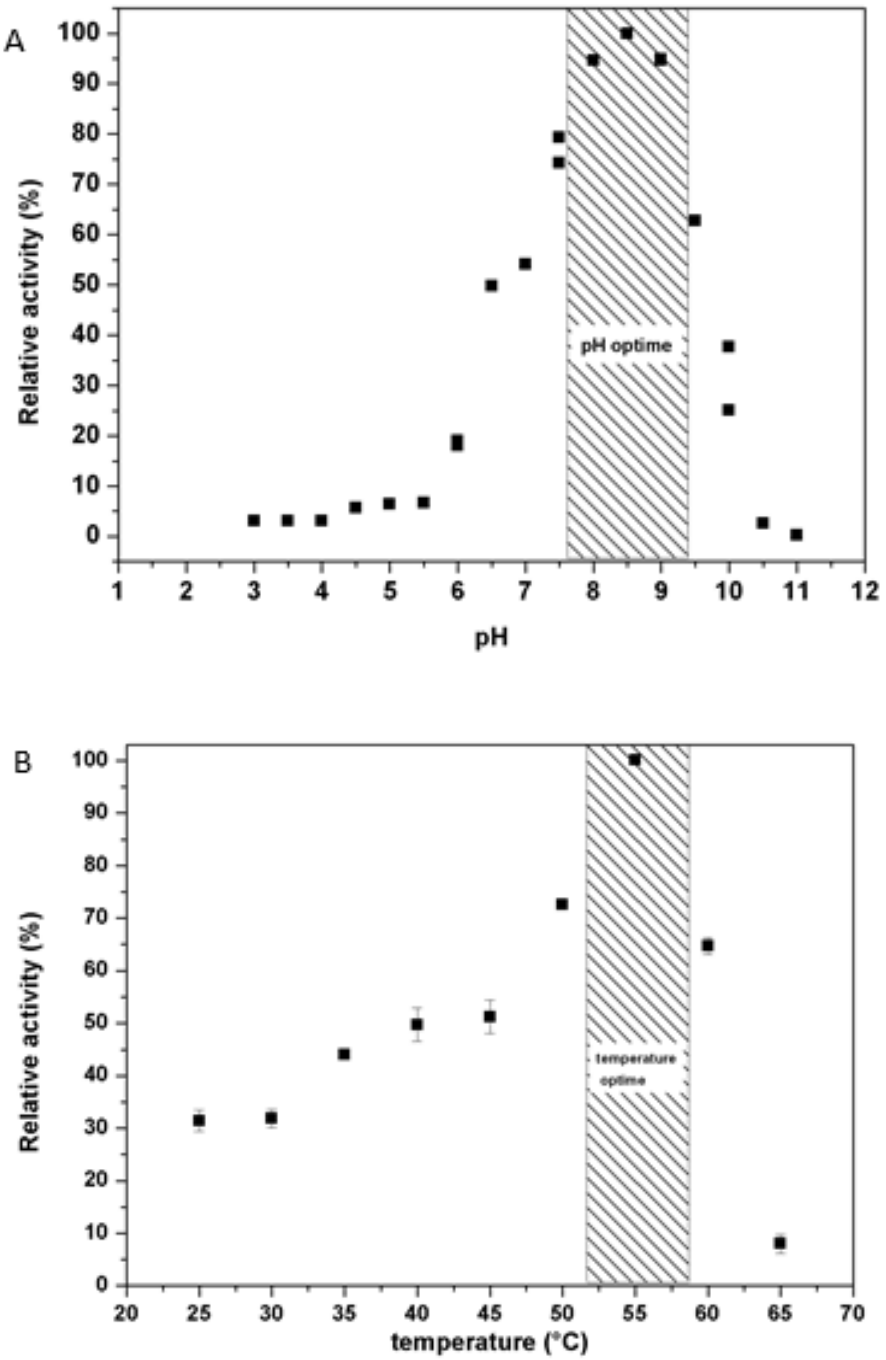


Figure 12 – Characterization of lipX.
A The effect of pH on enzyme activity was determined at 30°C using 50 mM of each of following buffers: KCl/HCl (pH 1–2.2), glycine/HCl (pH 2.2–3.6), sodium citrate (pH 3.0–6.0), sodium phosphate (pH 6.0–8.0), Tris/HCl (pH 7.6–9.0), CHES (pH 8.8–10.0), sodium carbonate (pH 9.6–10.6), and CAPS (pH 9.5–11). B The effect of temperature (25°C–50°C) was determined in a reaction mixture containing 1 mM pNP-butyrate and 150 nM LipX in PBS buffer (pH 7.5).

Enzyme activity of lipY was observed from pH 3.0–11.0, with optimal activity at pH 8.0–9.5 (Fig. 13A). Thus, lipY is not only active over a wide range of pH values but it is active under neutral as well as alkaline conditions. Within the temperature range of 25°C–95°C, maximum activity was observed at 60°C; relative enzyme activity was >60% from 50°C–55°C and approximately 50% at 65°C (Fig. 13 B).

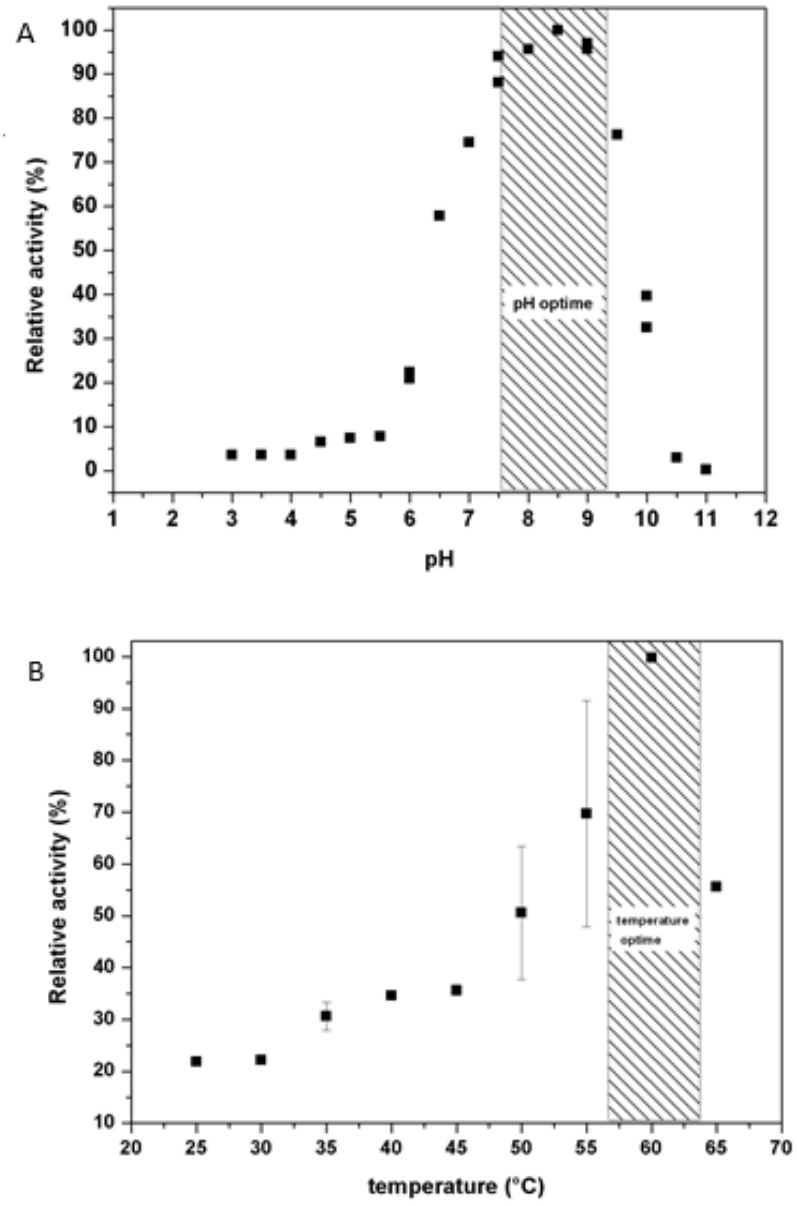


Figure 13 – Characterization of lipY
A The effect of pH on enzyme activity was determined at 30°C using 50 mM of each of following buffers: KCl/HCl (pH 1–2.2), glycine/HCl (pH 2.2–3.6), sodium citrate (pH 3.0–6.0), sodium phosphate (pH 6.0–8.0), Tris/HCl (pH 7.6–9.0), CHES (pH 8.8–10.0), sodium carbonate (pH 9.6–10.6), and CAPS (pH 9.5–11). B The effect of temperature (25°C–50°C) was determined in a reaction mixture containing 1 mM pNP-butyrate and 150 nM LipY in PBTx buffer (pH 7.5)

Lipolytic enzymes of the HSL family exhibit diverse optima for temperature and pH. lipX and lipY was derived from a Cerrado soil metagenome, and as such it exhibits maximum catalytic activity under mesophilic conditions (below 50°C). In addition, they display thermal tolerance at moderate temperatures. The ability for lipolytic enzymes to evolve thermal adaptations that resemble their bacterial host habitats. They did not lose activity below 30°C, a feature that highlights lipX and lipY as a candidate's catalyst in low temperature bioprocesses. An adaptation to low working temperatures is an attractive trait for reactions in which high temperatures are not suitable, such as manufacture of thermolabile pharmaceutical products, food ingredients as well as production of cold-wash detergents(Kovacic et al., 2015).

Structural analysis of l lipX and lipY by Fluorescence and CD

Fluorescence essays for LipX were performed at pHs 4.5 to 9.5. The tryptophan emission spectra showed a maximum fluorescence emission at 331 nm pH 4.5-6.5 and 333 nm at pH 7.0-pH 9.5 (Figure 14). This small shift of the emission band to higher wavelengths suggests changes in the ionic environment around tryptophan, which is positioned in a semi-buried region to regions more exposed to the solvent. Fluorescence emission spectra were pH-dependent, increasing the emission intensity between pH 5.0 -8.0 and 9.0-9.5. At pH 8.5 there was a decrease in the tryptophan emission band and displacement to 333 nm. The variation of pH-dependent fluorescence is influenced by the suppressor groups adjacent to tryptophan. Ionization of the side chains of acidic amino acid residues such as aspartic acid, glutamic acid, histidine, glycine and arginine which are close to tryptophan, can result in a suppressor group.

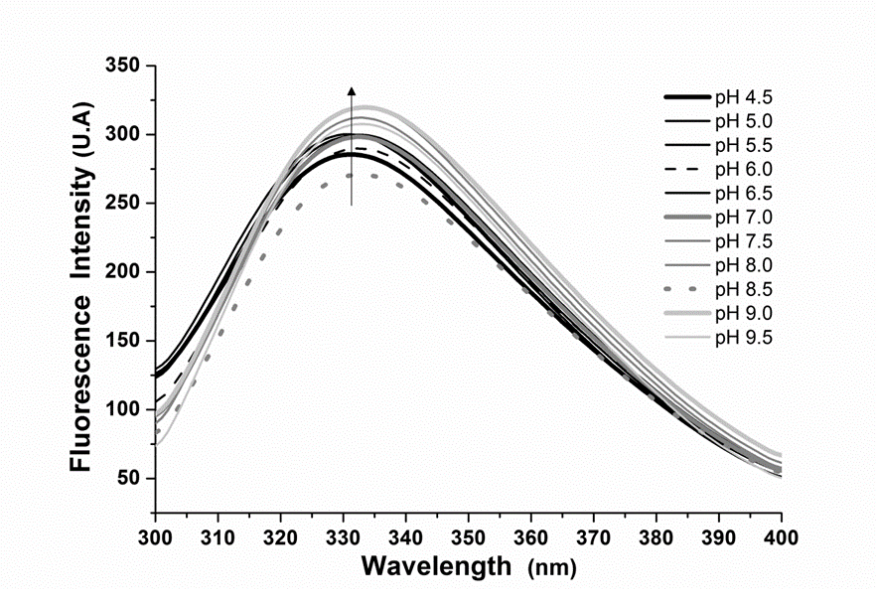


Figure 14 – Fluorescence spectra of lipX as a function to a broad pHs (4.5-9.5).
The fluorescence spectra were shifted by approximately 2 nm from wavelength from 331 at pHs 4.5-6.5 to 333 at pHs 7.0-9.5.

The secondary structure of the LipX enzyme was characterized by circular dichroism in the far UV region (190-260 nm) at 25 °C as a function of different pHs. Dichroic spectra have shown that the secondary structure of LipX is pH dependent. The spectra shows pronounced negative dichroic bands at 208, 218 and 222 nm and positive band at 195 nm, however, the signal is more evident at pH 7.0 (Fig15). This result shows that the protein is more structured in the neutral range (Table 2). The secondary structure content calculated for lipX at different pHs showed predominant structure in α-helix, at pH 7.0 the α-helix content was 38.3%; reducing by 6% and 2%, at pHs 5.0 and 8.5 respectively, followed by β-sheet 14.6%. The experimentally determined secondary structure contents are in agreement with 3D structure of α / β esterases and lipases deposited in the Protein Data Bank (4YPV, 1QZ3, 2HM7, 5JD4), which have α-helix centered secondary structure ~36%, followed by sheet β ~15%. The 3D homology model constructed for LipX is also in agreement with the experimental data obtained by CD.

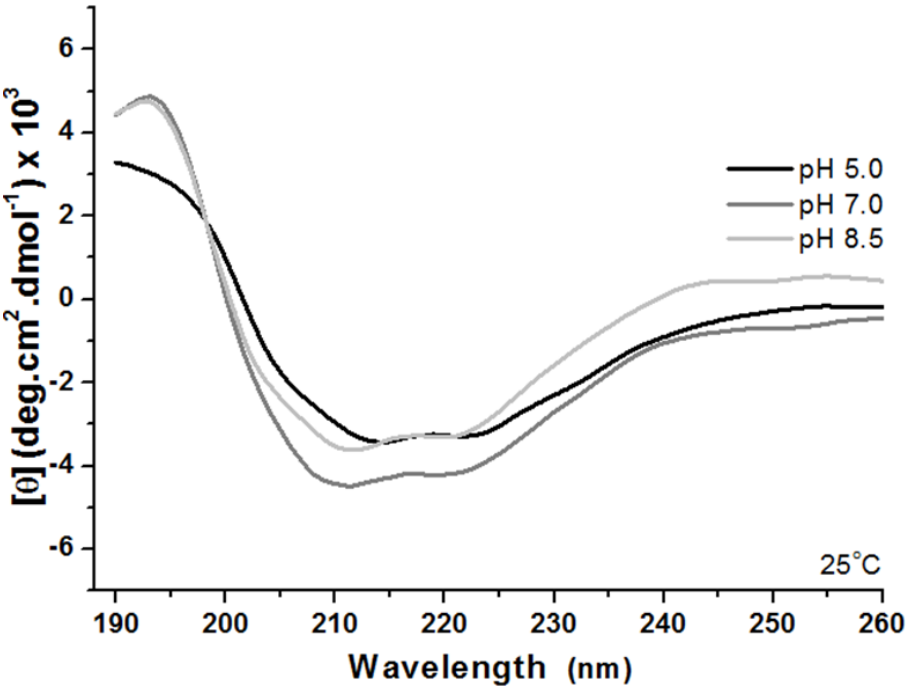


Figure 15 – Far-UV Circular Dichroism spectra of lipY as a function of pH at 25°C.
The LipX (0.128 mg/mL) was solubilized in 2 mM sodium acetate buffer pH 5.0 and 2 mM Tris HCl buffer pH 7.0 and 8.5.

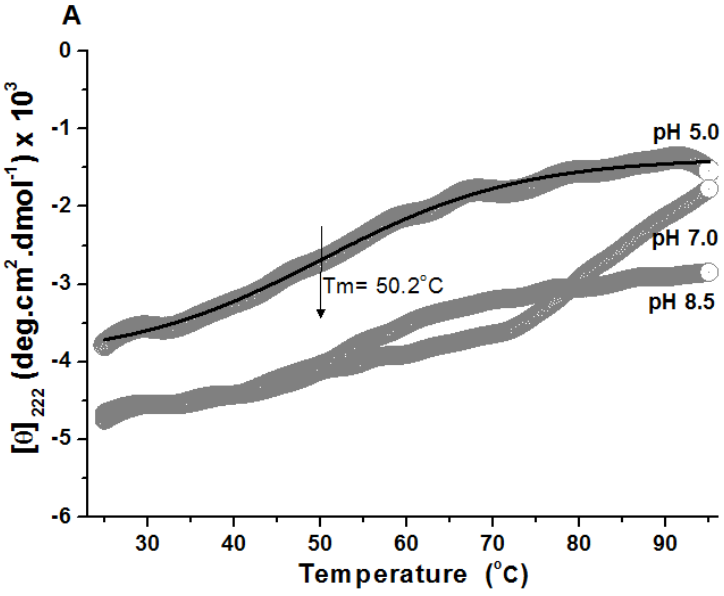
***Table 5** – The secondary structure contents of LipX at pH 5.0, 7.0 and 8.5 obtained by the CDNN deconvolution software*

Secondary structure (%)	pH 5.0	pH 7.0	pH 8.5
Alpha helix	32.5	38.3	36.2
β-antiparallel	8.0	9.1	7.3
β-parallel	9.3	5.5	8.2
Turn-β	16.8	16.8	16.3
Random coil	34.1	28.1	31.2

The thermal stability of the LipX enzyme was characterized at different pHs.

Thermal denaturation curves indicated partial or total denaturation of the secondary structure of LipX protein as a function of temperature (Figure 16 A-D). The dichroic spectra collected in the denaturation (195-260 nm) essays at pHs 5.0, 7.0 and 8.5 showed a decrease in the 208, 218 and 222 nm negative bands. The denaturation curves (λ222 nm) for pHs 5.0 and 7.0 showed a gradual decrease of the signal varying between -

4,500 and -1,200 and the spectra collected concomitantly showed the loss of the native structure as a function of the temperature increase, since there was an evident decrease of the bands corresponding to the α-helix structures (208 and 222 nm) and β-sheets (195 and 218 nm), suggesting complete denaturation of the protein, this process was not reversible for both pHs. The denaturation process for the protein at pH 7.0 showed state of aggregation at high temperatures, so it was possible to calculate the transition temperature between the native state and unraveled only at pH 5.0 Tm = 50.2 ° C. However, for the protein at pH 8.5, denaturation occurred partially, with maintenance of the dichroic signal from -4,500 to -2,500 at 222 nm and partial reduction of the dichroic signal from the spectra at the bands regions corresponding to the α-helix structures (208 and 222 nm) and β sheets (195 and 218 nm).



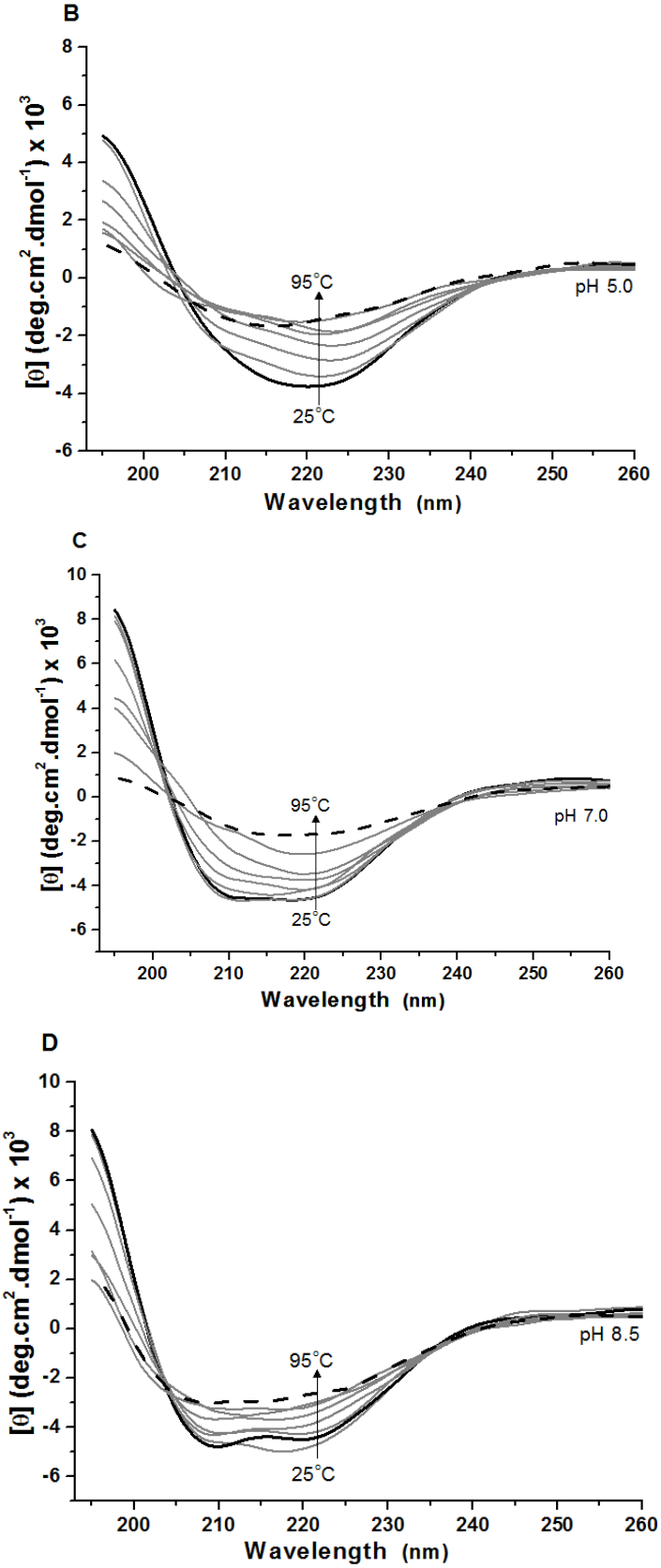


Figure 16 – Thermal stability analysis of the lipX at pHs 5.0, 7.0 and 8.5 as a function of temperature. *A) The unfolding curves of the LipX enzyme (0.128 mg/mL) were recorded at λ222 nm, varying the temperature from 25 to 95°C, using sodium acetate (2 mM) and Tris HCl (2 mM). The black lines*

correspond to the sigmoid fitting of experimental data. Tm are indicated by vertical arrows. B-D) Far-UV CD spectra of LipX by increase temperature from 25 to 95oC indicating reduction of the molar ellipticity [θ] (arrows)

The results clearly showed that lipX is more efficiently in basic pH (8.0 and 9.0) relative to neutral pH. On the other hand, the loss of activity was not due to differences in the secondary structure of lipX, which reveals an alpha-beta structure, as is expected for esterases/lipases enzymes. This result is in agreement with the structural model and with the CD analysis that revealed negative peaks of 210 nm (shift from 208 nm) and 222 nm, characteristics of alpha-beta secondary structure content.

Fluorescence essays for lipY were performed at pHs 4.5 to 9.5. The emission spectra of tryptophan showed a maximum fluorescence emission at 338 nm for pH 4.0 and 336 nm for the other pHs analyzed (Figure 17). This small shift of the emission band to smaller wavelengths suggests changes in the ionic environment around the tryptophan, which is positioned in regions more exposed to the solvent. The fluorescence emission spectra were pH-dependent, increasing the emission intensity from pH 4.5 to 5.5, pH 6.5 to 9.0. At pHs 6.0 and 9.5 there was a decrease in the tryptophan emission band. The variation of pH-dependent fluorescence is influenced by the suppressor groups adjacent to tryptophan. Ionization of the side chains of acidic amino acid residues close to tryptophan, can result in a suppressor group.

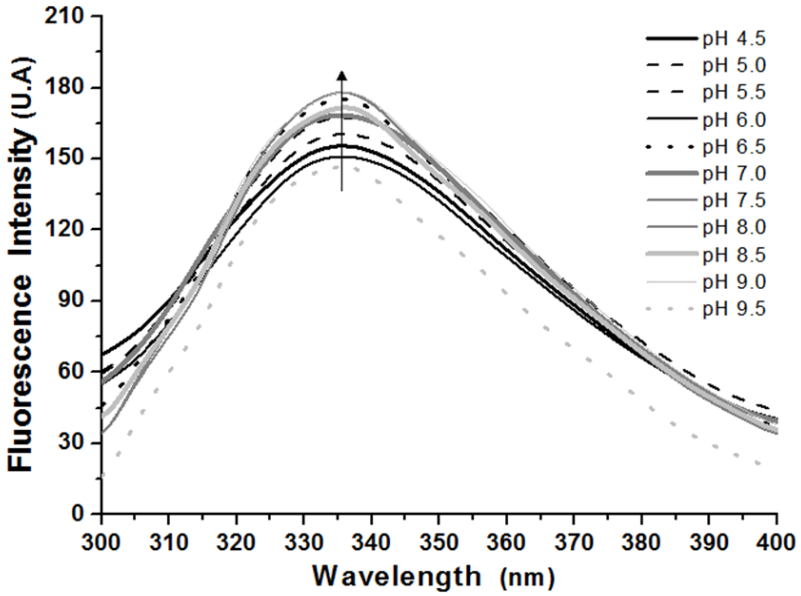


Figure 17 – Fluorescence spectra of lipY as a function to a broad pHs (4.5-9.5).
The fluorescence spectra were shifted by approximately 2 nm from wavelength 338 at 336 nm, from pHs 4.5 at at pH 5.0-9.5.

The secondary structure of the LipY enzyme was analyzed by circular dichroism in the far UV region (190-260 nm) at 25 ° C as a function of different pH. The collected dichroic spectra has shown that the secondary structure of LipY is pH dependent. Negative dichroic bands at 208, 218 and 222 nm showed a small reduction at pH 5.0 and 8.5 and increase of the positive band at 195 nm respectively (Figure18).

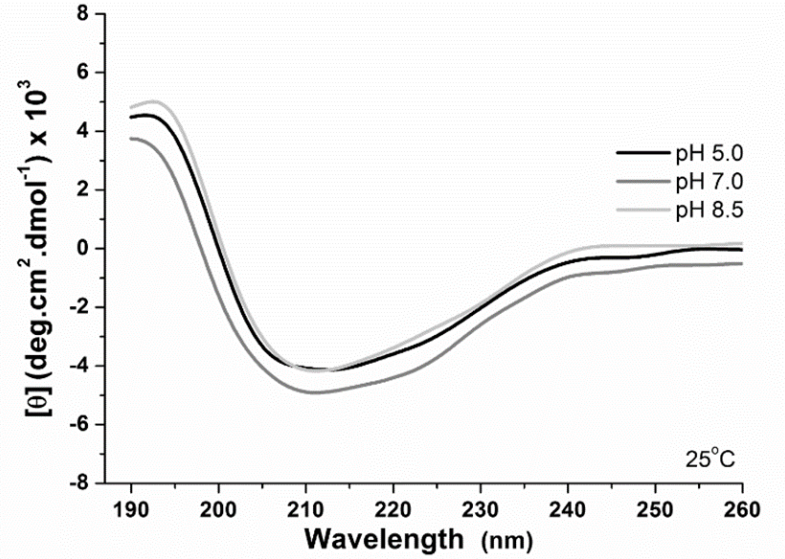


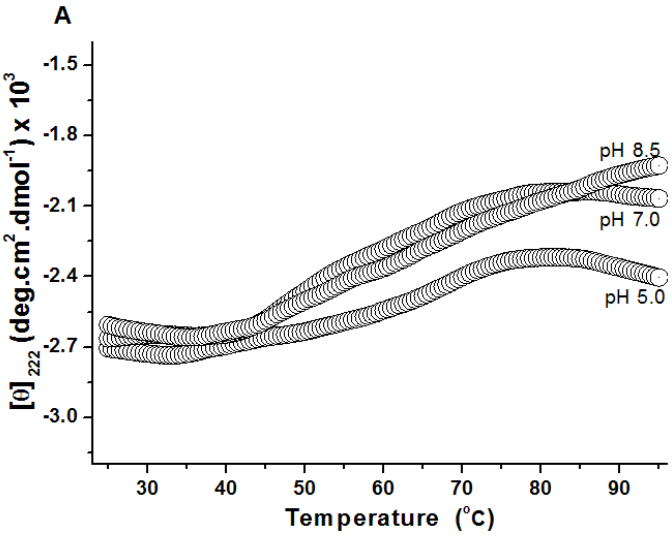
Figure 18 – Far-UV Circular Dichroism spectra of lipY as a function of pH at 25°C.
The LipY (0.125 mg/mL) was solubilized in 2 mM sodium acetate buffer pH 5.0 and 2 mM Tris HCl buffer pH 7.0 and 8.5.

This result shows that the protein is more structured in the alkaline range, with a molar theta around -4.200 and compliance with calculated secondary structure content (Table 3). The secondary structure content calculated for LipY at different pHs had a predominant α-helix structure, at pH 7.0 the α-helix content was 39.7%; reducing by 5% and 8%, at pHs 5.0 and 8.5 respectively, followed by β ~ 15% sheet. The experimentally determined secondary structure contents are in agreement with 3D structure of esterases and lipases deposited in the Protein Data Bank (5GMS, 4XVC, 3V9A), which have α-helix centered secondary structure ~ 36%, followed by β ~ 15%. The 3D homology model constructed for LipY is also in agreement with the experimental data obtained by CD.

Table 6 *– The secondary structure contents of LipY at pH 5.0, 7.0 and 8.5 obtained by the CDNN deconvolution software*

Secondary structure (%)	pH 5.0	pH 7.0	pH 8.5
Alpha helix	34.1	39.7	31.7
β-antiparallel	11.5	7.9	11.5
β-parallel	5.6	5.3	5.5
Turn-β	17.2	17.6	19.8
Random coil	30.6	27.2	31.3

The thermal stability of lipY was characterized at different pHs. Thermal denaturation curves has indicated alterations in the secondary structure of the protein as a function of pH and temperature increase (Figure 19 A-D). The dichroic spectra collected in the denaturation (195-260 nm) essays at pHs 5.0, 7.0 and 8.5 has showed a slight decrease in the 222 nm negative band ranging from -2,700 to -2,000 and a major change in the 195 nm positive band ranging from 5000 for 2000 as a function of temperature, suggesting partial denaturation of the protein. This shows that the LipY enzyme is thermostable by varying the temperature. This result is in agreement with the data obtained in the biochemical characterization, where the enzyme presented an optimum temperature of 60 ° C.



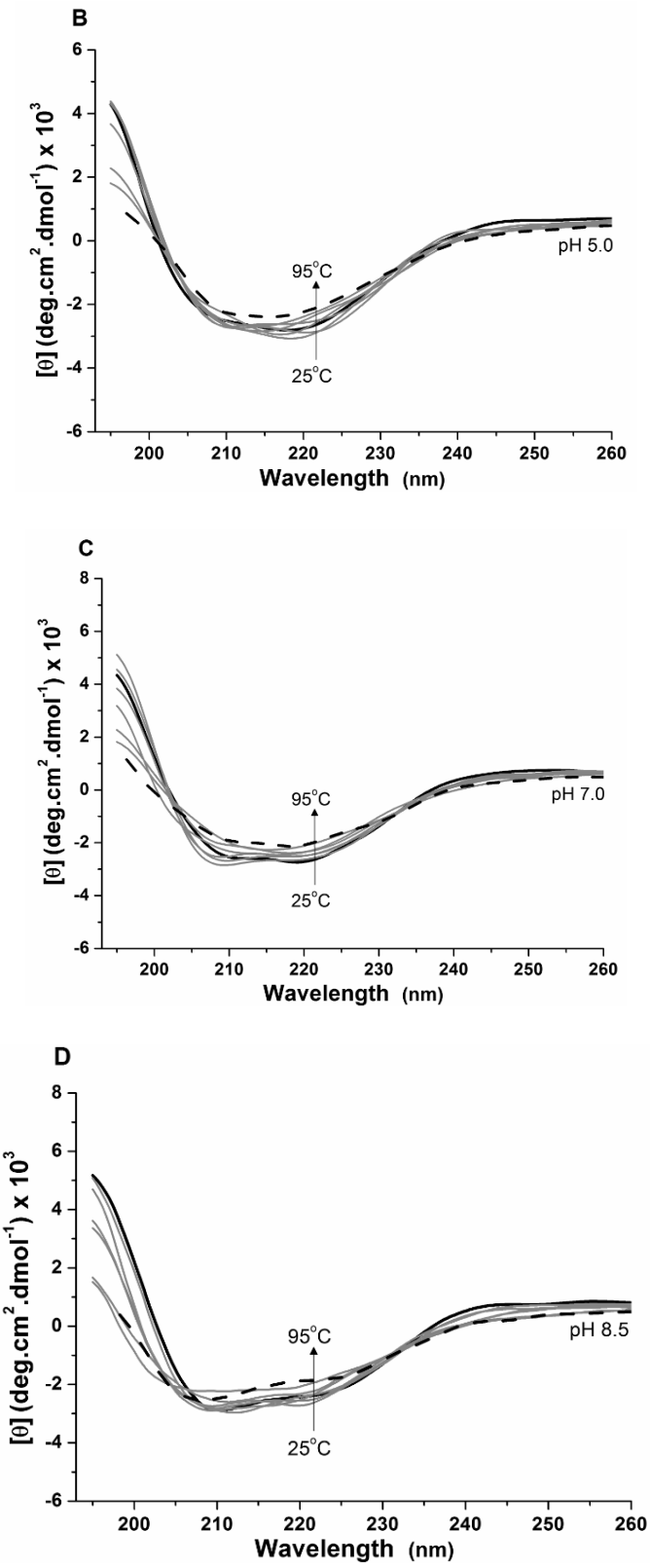


Figure 19 – Thermal stability analysis of the lipY at pHs 5.0, 7.0 and 8.5 as a function of temperature. *A) The unfolding curves of the LipY enzyme (0.125 mg/mL) were recorded at λ222 nm, varying the temperature from 25 to 95°C, using sodium acetate (2 mM) and Tris HCl (2 mM). The black lines correspond to the sigmoid fitting of experimental data. Tm are indicated by vertical arrows. B-D) Far-UV CD spectra of LipY by increase temperature from 25 to 95oC indicating small reduction of the molar ellipticity [θ] (arrows).*

The results clearly showed that lipY is more efficiently in basic pH (8.0 and 9.0) relative to neutral pH. On the other hand, the loss of activity was not due to differences in the secondary structure of lipY, which reveals an alpha-beta structure, as is expected for esterases/lipases enzymes. This result is in agreement with the structural model and with the CD analysis that revealed negative peaks of 210 nm (shift from 208 nm) and 222 nm, characteristics of alpha-beta secondary structure content.

Structural Analysis of lipX and lipY

The results from the comparation of the lipX with BLAST algorithm against protein data bank (PDB) exhibited 41% identity with G84S EST2 mutant from *Alicyclobacillus acidocaldarius* [PDB 2HM7]. The coordinate of the structure of 2HM7 were used for lipX model building based on similar function and amino acid sequence identity. The comparation resulted into a predicted LipX structure characterized by the α/βhydrolase fold from the IV Esterase/Lipase hormone sensitive Family, with the catalytic residues: Ser156, Asp 259 and His 289 located between α/β domain and the helical domain (Fig 20 and 21). The Ser 156 is located in the consensus motif: GDSAGG, as also showed in the sequence alignment of lipX with other members of related Family IV.

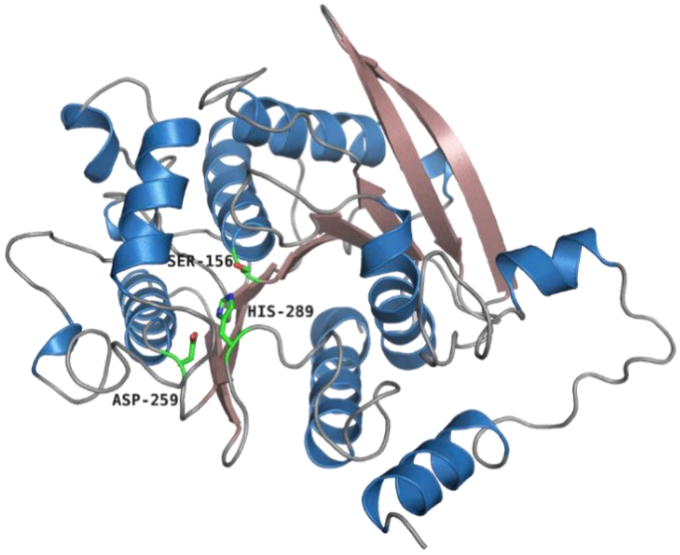


Figure 20 – The modeler three-dimensional structure of lipX, 328 aa.
Catalytic residues: forming the triad are Ser156, Asp 259 and His 289 indicated in stick representation.

The structural comparison between LipX and 2HM7 in the region of the active site reveals that the active site, formed by the catalytic triad and the oxyanion hole is well conserved.

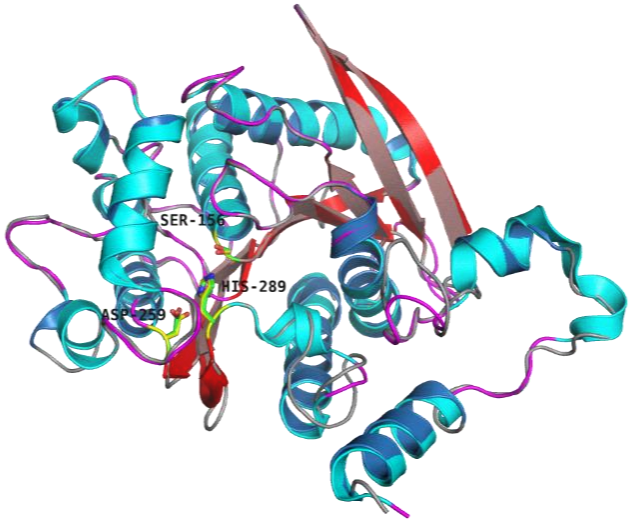
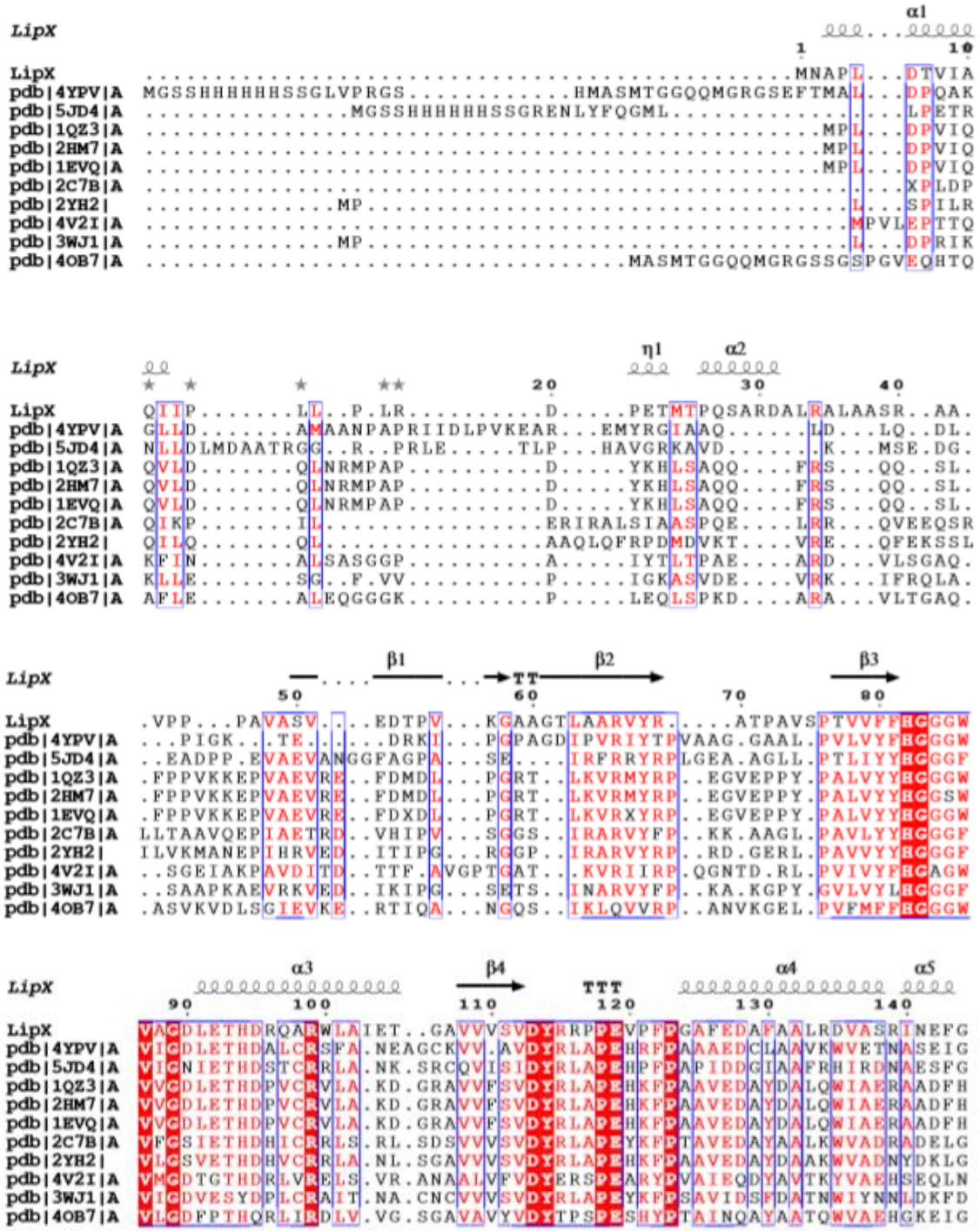


Figure 21 – Superposition of the modelled lipX structure onto the structure of the 2HM7 protein with 41% identity. RMS = 0. 094.*The active sites of each enzyme are located between the two domains and the superposition revealed that the residues from the catalytic triad are located exactly in the same positions*

Other characteristic of Esterase/Lipase hormone sensitive family, the formation of the oxyanion hole of lipX located in the HGG motif is constituted by residues G83, G84, whose main chain nitrogen atoms act as hydrogen donors to the cleaved substrate, stabilizing the negative charge on the tetrahedral intermediates arising from the nucleophilic attack of the Ser156 from the catalytic triad. The multiple sequence alignment of LipX made with besthit esterases of known 3Dstructure display the conserved amino acids highlighted in red. Elements of secondary structure for all proteins are denoted as α (α Helix), β (β sheet), η (random coil), T (β turn) (Figure 22).


LipX
α1
1
10
LipX
pdb|4YPV|A
pdb|5JD4|A
pdb|1QZ3|A
pdb|2HM7|A
pdb|1EVQ|A
pdb|2C7B|A
pdb|2YH2|
pdb|4V2I|A
pdb|3WJ1|A
pdb|4OB7|A
η1
α2
20
30
40
β1
β2
β3
50
60
70
80
α3
β4
α4
α5
90
100
110
120
130
140

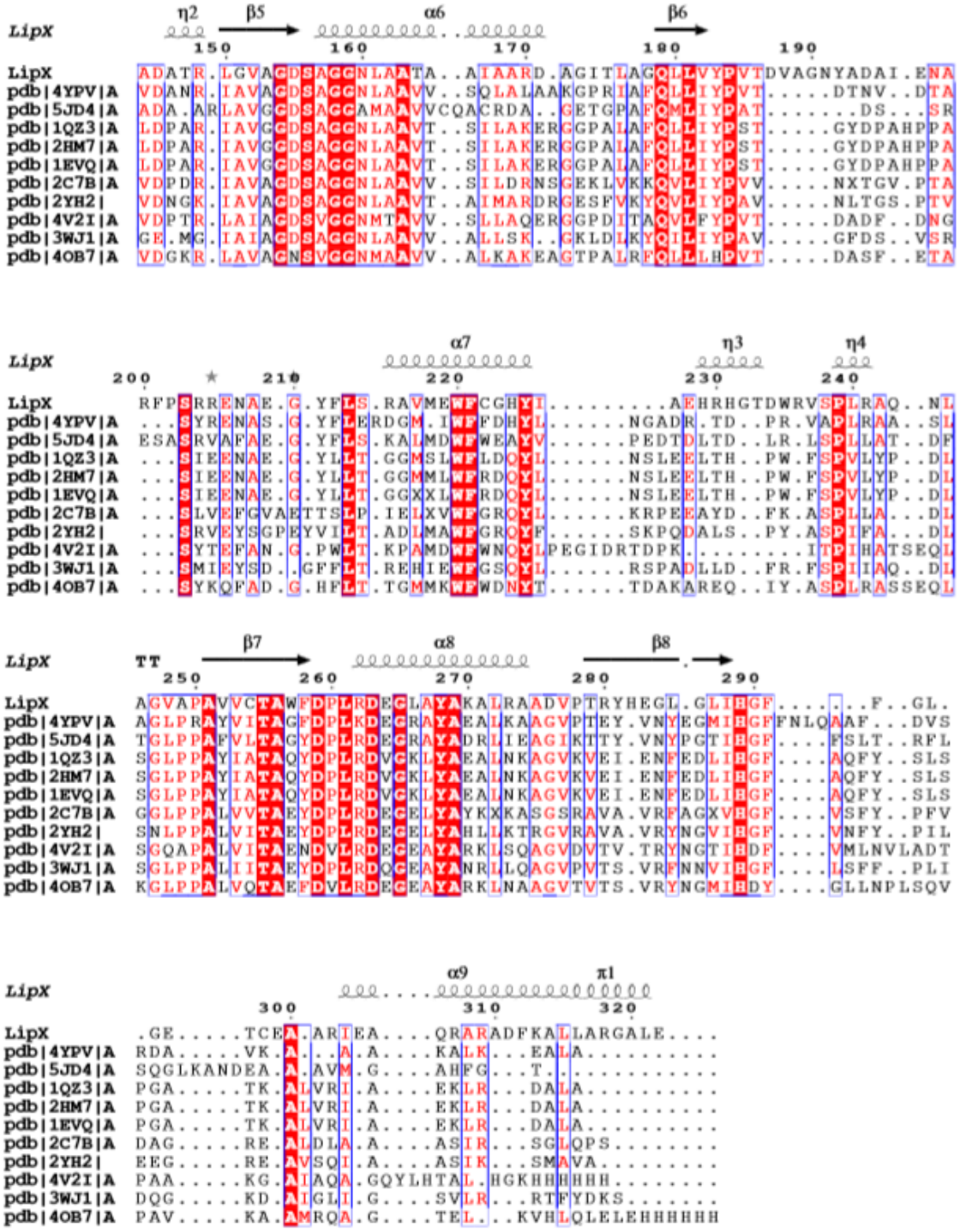


Figure 22 – Multiple sequence alignment of lipX and besthits esterases of known 3Dstructure.
The conserved amino acids are highlighted in red. Elements of secondary structure for all proteins are denoted as α (α Helix), β (β sheet), η (random coil), T (β turn).

The stereochemical quality of the theoretical model was evaluated by the construction of the Ramachandran chart. The model of lipX presented 92.2% of the amino acids located in favorable positions, 5.6% in allowed regions and 2.2% in non-permitted regions, implying that the model presented good stereochemical quality (Fig 23 A).

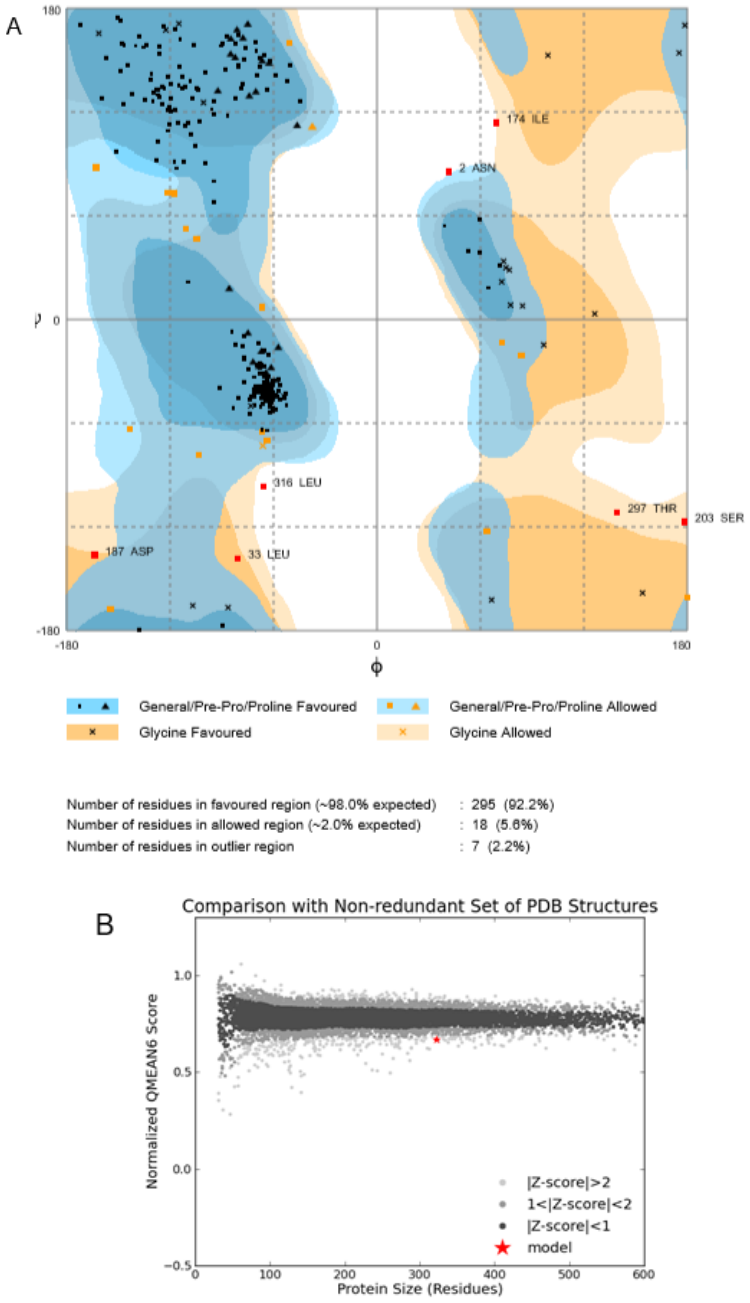


Figure 23 – Evaluation of the theoretical model constructed by comparative molecular modeling of lipX *A) Ramachandran plot of the final model of lipX obtained by modeling molecular theory.: B) Absolute quality chart estimated by QMEAN Z-score for the theoretical model of lipX*

The multiple sequence alignment of lipY with other esterases that had the highest identity in the PDB database shows that lipY has a catalytic triad formed by Ser143, Glu237, and His267. LipY also contains a CAP domain (Met1–Val45) beside catalytic domain (Gln46–Arg306) (Figure 24). The catalytic Ser143 is located in the semi-conserved GXSAG motif where instead of Asp (GDSAG), the most common; there is another negative charged residue, Glu (GESAG) (Li et al. 2015). Like other HSLs, lipY contains a conserved HGG motif. This microbial motif is involved in the formation of the oxyanion hole, and it is in close proximity to the catalytic triad (Mandrich et al. 2008). Residues Gly within the conserved HGG motif comprise the oxyanion hole that is involved in substrate binding for HSL esterases (Li et al. 2015) (Figure 25).

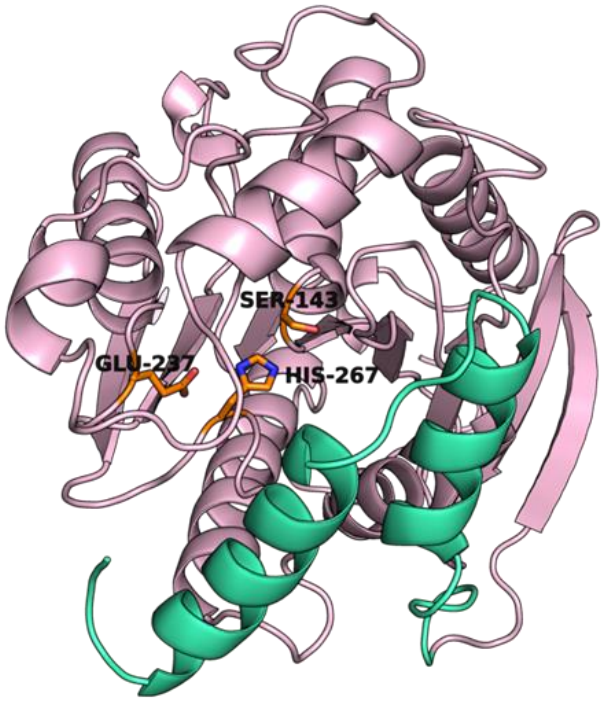


Figure 24 – Structural features of the lipY model.
The structure of the LipY are shown in the cartoon. The CAP domain is shown in green, and the catalytic domain is in pink.

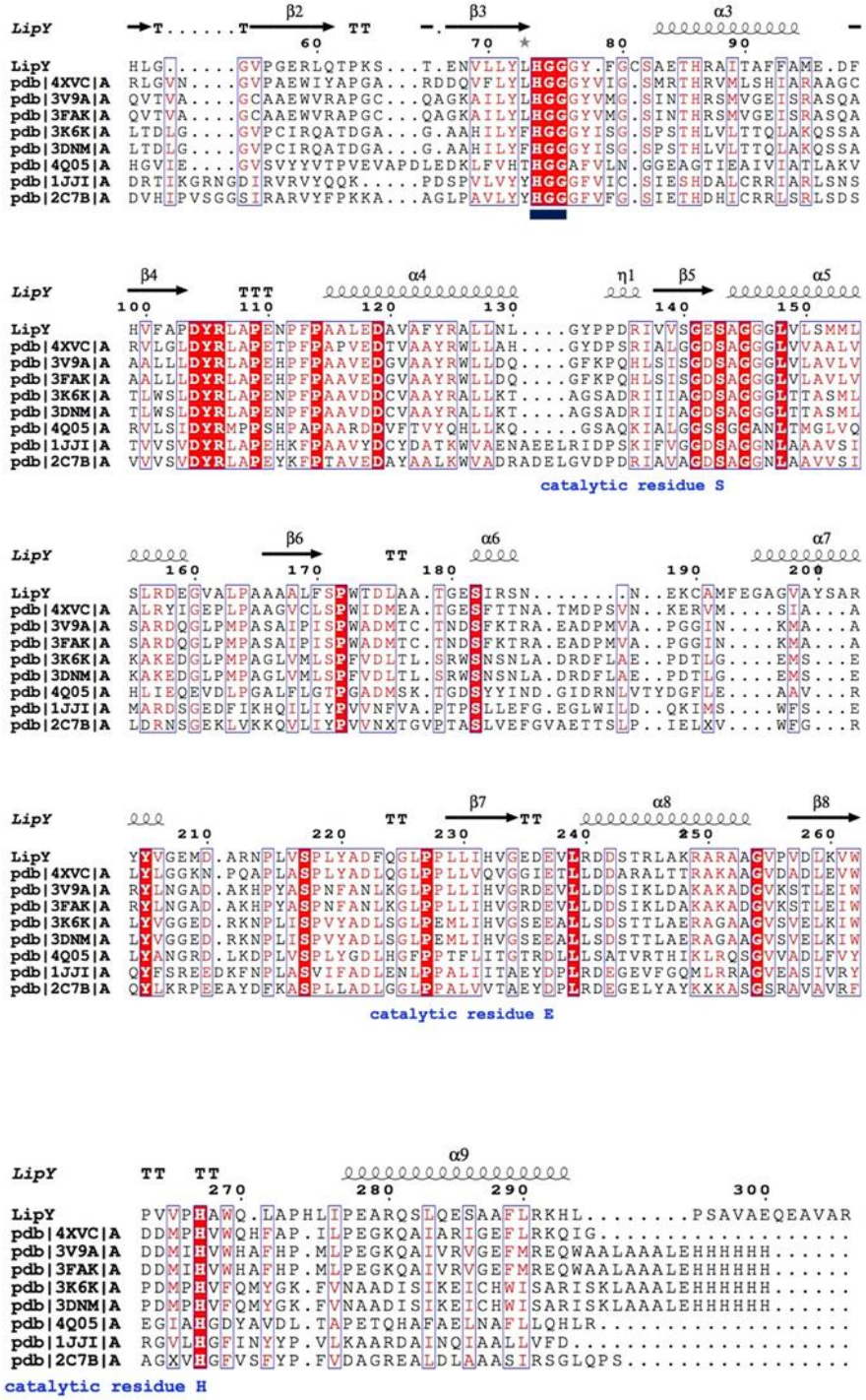


Figure 25 – The amino acid sequence alignment of lipY with other esterases from the GDSAG motif subfamily of the HSL family. *The residues involved in the oxyanion hole (HGG) are shown conserved. The catalytic triad residues are shown semi-conserved. α: α-helix, β: β-sheet, e T: β-turn.*

The stereochemical quality of the theoretical model was evaluated by the construction of the Ramachandran chart. The model of LipY presented 92.4% of the amino acids located in favorable positions, 4.9% in allowed regions and 2.6% in non-permitted regions (Fig 26 A).

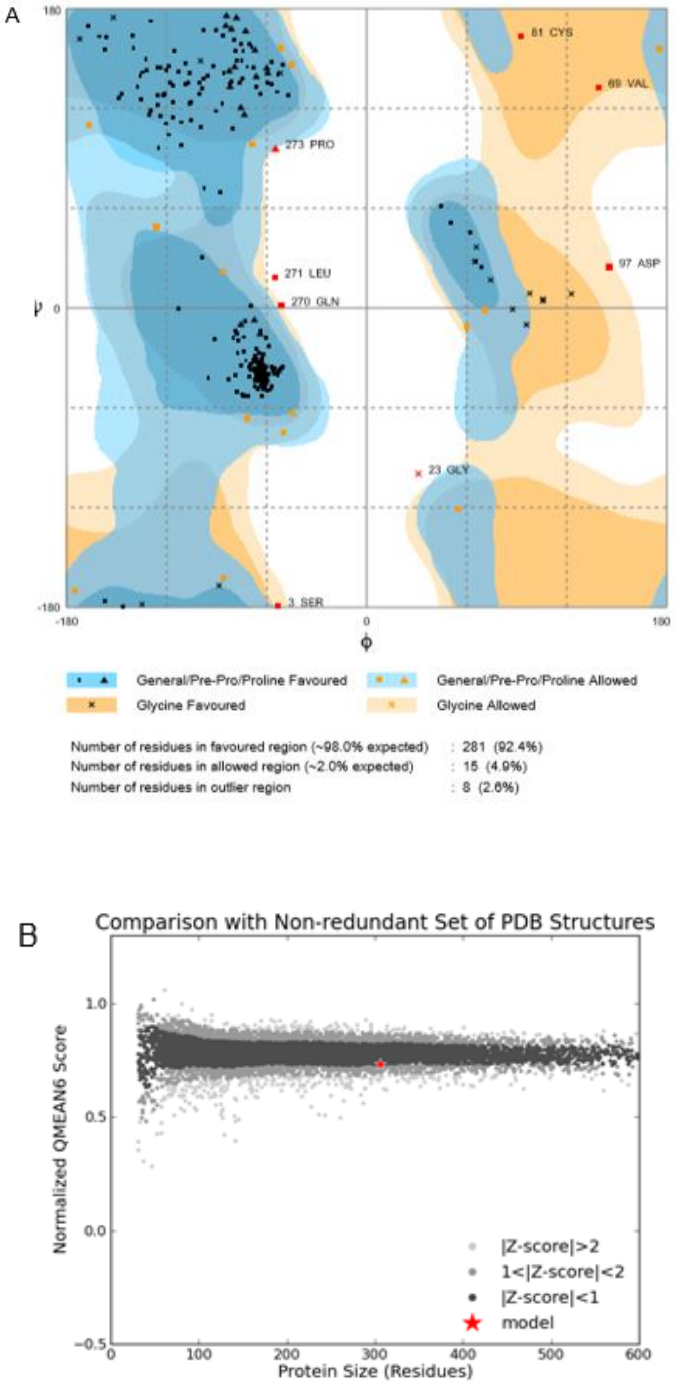


Figure 26 – Evaluation of the theoretical model constructed by comparative molecular modeling of lipY: *A) Ramachandran plot of the final model of lipX obtained by modeling molecular theory.: B) Absolute quality chart estimated by QMEAN Z-score for the theoretical model of lipX*

To understand the structural basis for lipY thermolability, the model was compared with two HSL esterases: E40 (thermolabile esterase) and Est2 (thermophilic esterase). Hydrophobic interactions between the loop 1 localized in the CAP domain and the loop 7 of the catalytic domain seems to be related to the thermal stability of microbial HSLs (Li et al. 2015). Compared with E40, the distance between these loops is much closer than in Est2, suggesting that there may be less interactions between the CAP domain and the catalytic domain in LipY/E40 than in Est2 (Figure 27).

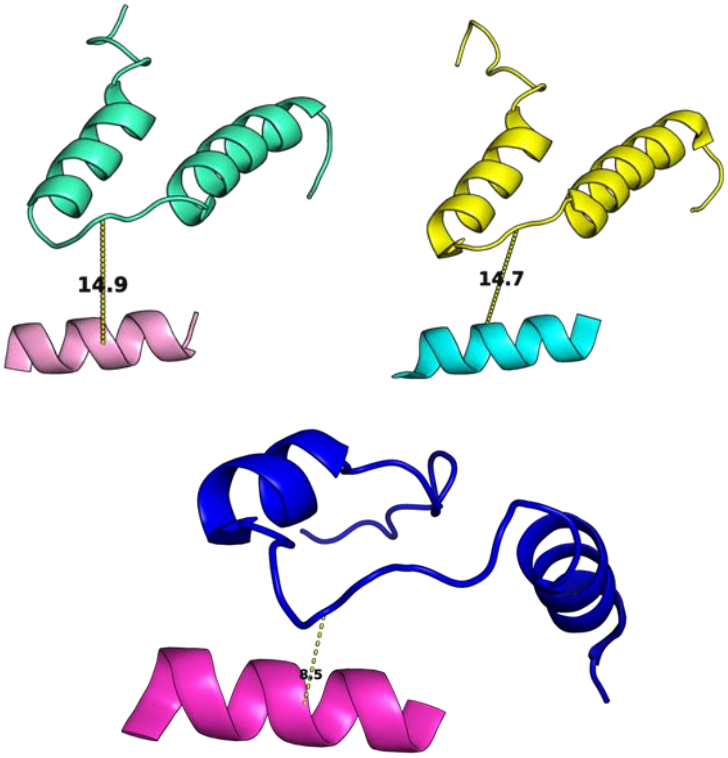


Figure 27 – Comparative structural analysis of lipY with E40 (thermolabile mesophilic) and Est2 (thermophilic).*The minimum distances between the main chain atoms of loop 1 (CAP domain) and 7 (Catalytic domain) are also shown for LipY (A), E40 (B) e Est2 (C). The distances are given in Angstrom.*

For thermophilic Est2, the loop 1 of the CAP domain forms interdomain hydrophobic interactions with α7 of the catalytic domain. These interactions occur among hydrophobic residues with large side chains present in the loop 1 and two adjacent hydrophobic residues: Trp and Phe in the α7 of the catalytic domain are conserved in thermostable HSLs. The secondary structure and multiple amino acid sequence alignment of LipY presented in the figure 25 shows that in the α7, lipY does not have Trp and Phe (Li et al. 2015). Besides that, the highest identity of the lipY was

with E40, suggesting that less interdomain interaction may occurs characterizing LipY as a possible thermolabile esterase. Further studies must be performed such as mutations followed by enzymatic activity to check this information.

The structural comparison between lipY and 4XVC in the region of the active site reveals that the active site, formed by the catalytic triad and the oxyanion hole is well conserved and 46% identity.

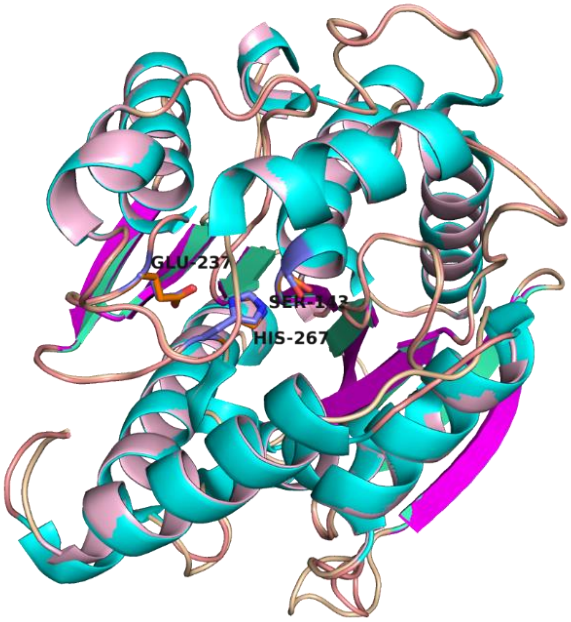
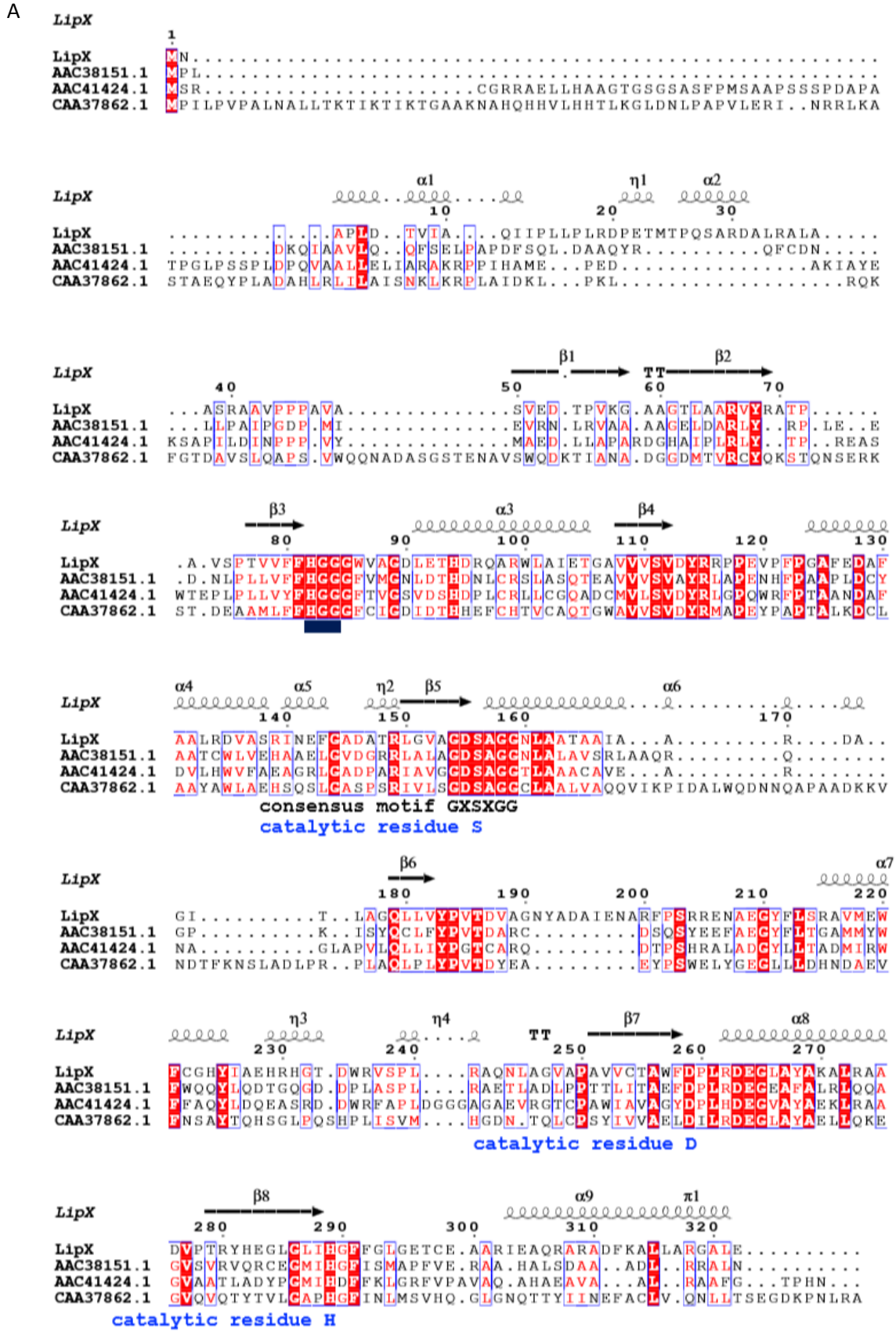


Figure 28 – Superposition of the modelled LipY structure onto the structure of the 4XVC protein with 46% identity.

The secondary structure and multiple amino acid sequence alignment of lipX with lipY shows that even if they came from the same ambient sample, they have distinct characteristics. Hormone-sensitive lipases (HSLs) exist widely in microorganisms, plants, and animals. Microbial HSLs are classified into two subfamilies: the GTSAG motif subfamily and the GDSAG motif subfamily. LipX is from GDSAG motif subfamily but lipY where instead of Asp (GDSAG), the most common; there is another negative charged residue, Glu (GESAG).

A


LipX
AAC38151.1
AAC41424.1
CAA37862.1
consensus motif GXSXGG
catalytic residue S
catalytic residue D
catalytic residue H

B

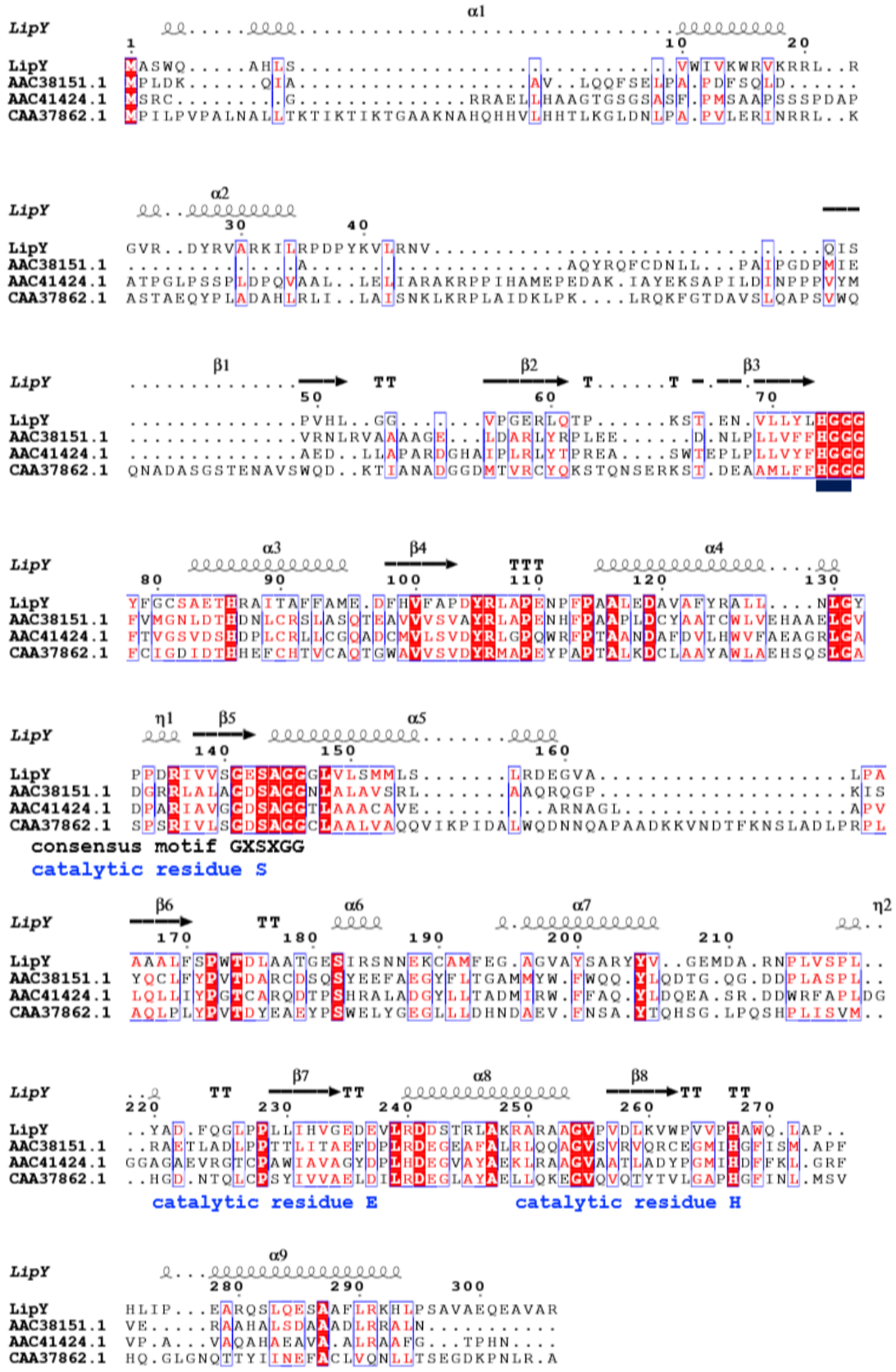


Figure 29 – Secondary structure comparation lipX and lipY

The amino acid sequence alignment of lipX and lipY with other esterases from HSL family. The residues involved in the oxyanion hole (HGG) are shown conserved. The catalytic triad residues are shown semi-conserved. α: α-helix, β: β-sheet, e T: β-turn.

GENERAL CONCLUSIONS AND PERSPECTIVES OF THIS WORK

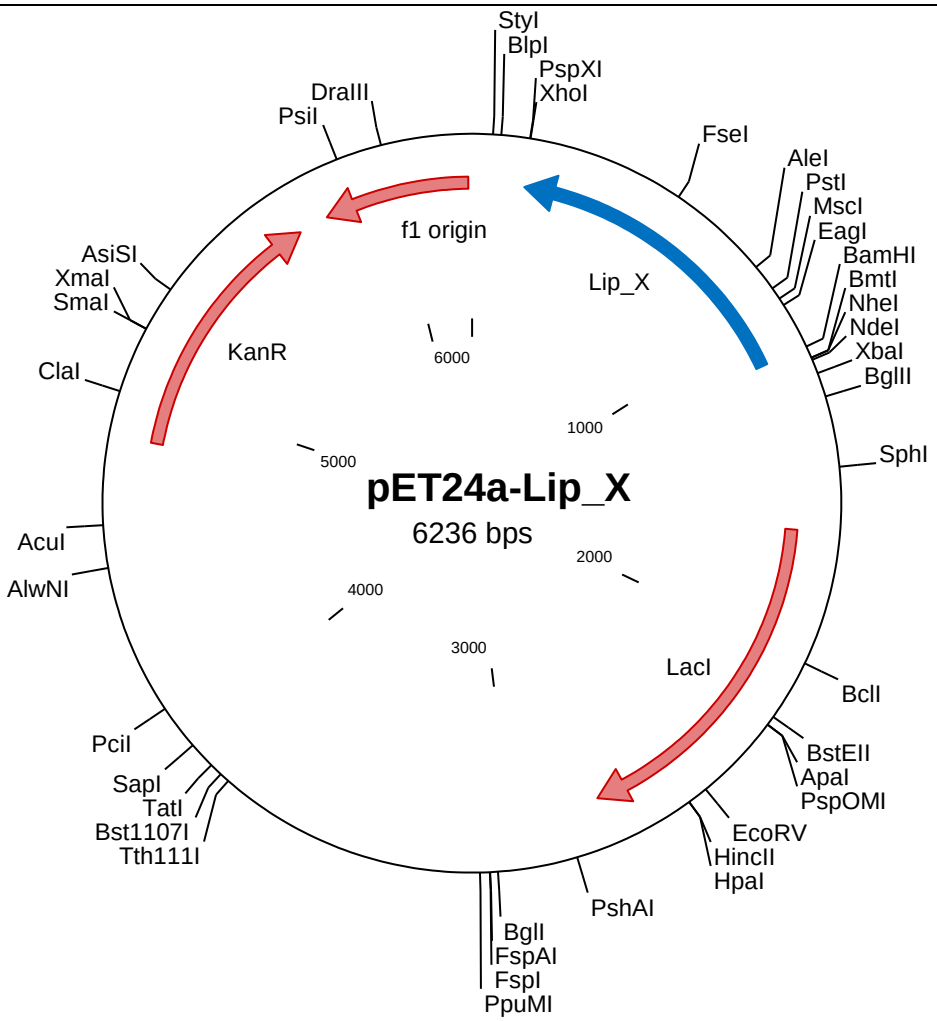
Three genes were identified as putative lipases/esterases and their activities have been further characterized in this work. The isolated and characterized enzyme lipX and lipY have showed conservation of the sequence and structural features of members from Family IV of bacterial lipolytic enzymes. The results from the comparison of the lipX with BLAST algorithm against protein data bank (PDB) exhibited 41% identity with G84S EST2 mutant from Alicyclobacillus acidocaldarius [PDB 2HM7] and α/βhydrolase fold from the IV Esterase/Lipase hormone sensitive family and lipW showed conservation of the sequence and structural features of members from family V of bacterial lipolytic enzymes. LipY also contains a CAP domain (Met1–Val45) beside catalytic domain (Gln46–Arg306). UV CD spectra and 3D model structure based on homology modeling indicate that lipX, Y and W secondary structure consisted of α-helix (~45%) and β-sheet (~15%) structures, which is consistent with the predominantly α-helix secondary structure of the α/β hydrolase family of esterase's. LipX and lipY contain a conserved HGG motif, this microbial motif is involved in the formation of the oxyanion hole.

The purification of lipX revealed a 35-kDa recombinant protein and lipY revealed a 32-kDa recombinant protein as predicted by the sequence data. The estimated molecular weight of LipW was 29 kDa. Molecular parameters could be associated to the lipW monomer in correspondence with their theoretical MW, as observed in the SDS-PAGE gel. Enzyme activity was observed from pH 3.0–10.0, with optimal activity at pH 7.5.0–9.5 for lipX, pH 8.0–9.5 for lipY and an optimal enzyme activity of lipW at pH 9.0–9.5. Within the temperature range of 25°C–95°C, maximum activity was observed at 55°C for lipX, 60°C for lipY and 40°C for lipW.

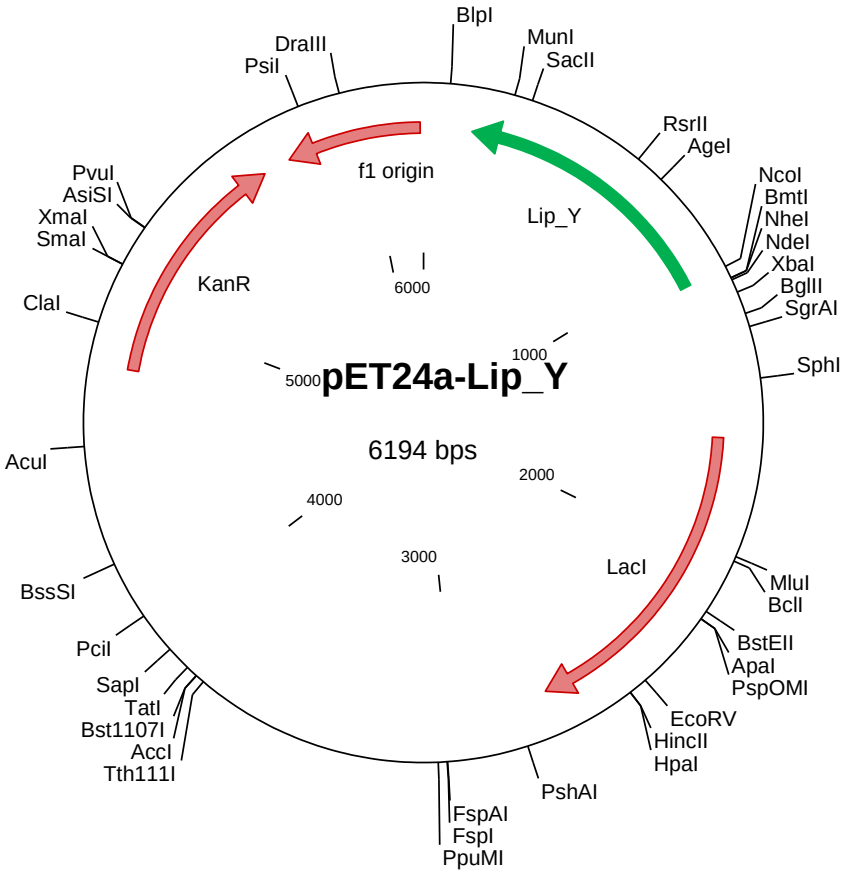
As perspectives from this thesis, cloning into X33/ Komagataella pastoris stain of lipx, lipY and lipW, large-scale production in fermenters and application tests with lipX, lipY and lipW separately or together making a cocktail for industry use.

A large amount of information that can be accessed with metagenomic approach can be used to cross-data, obtaining numerous information relevant to advances in biological and technological research, but also the integration of these metadata with the bioengineering of processes for a real application of them.

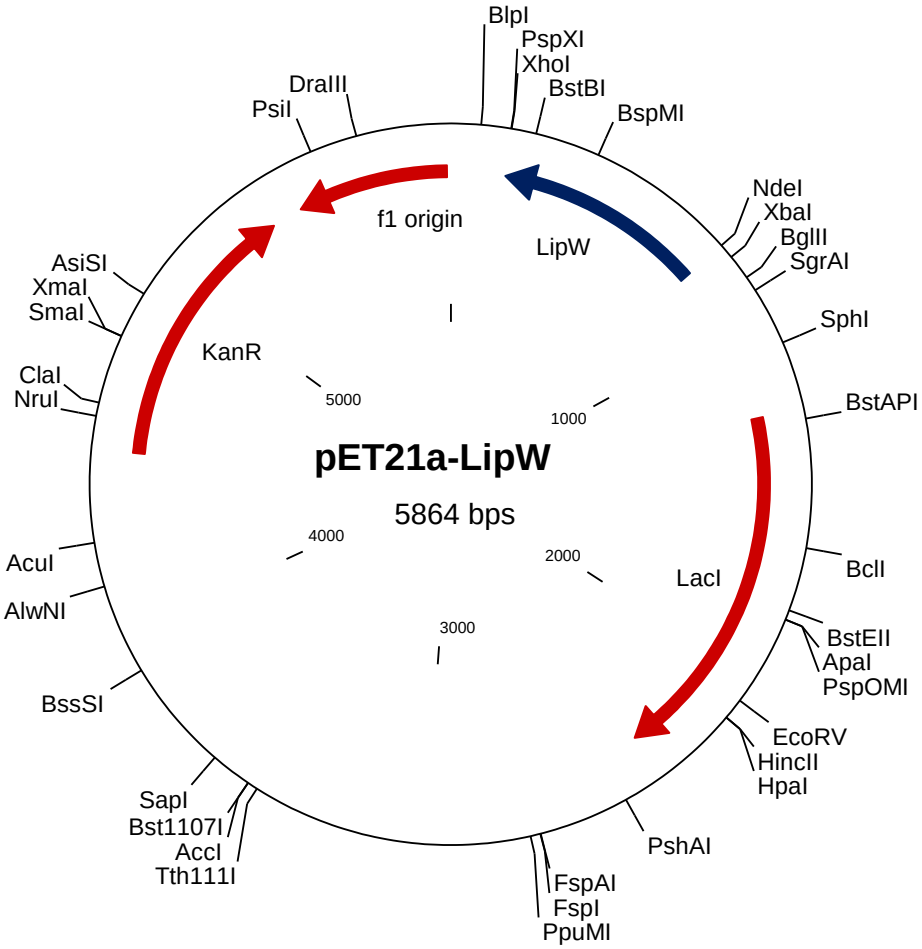
APPENDIX I



Construction of expression vector using pET24a (Novagen, USA), which carries an N-terminal His tag, to generate pET24a-lipX. with 924-bp fragment



Construction of expression vector using pET24a (Novagen, USA), which carries an N-terminal His tag, to generate pET24a-lipY. with 966-bp fragment



Construction of expression vector using pET21a (Novagen, USA), which carries an N-terminal His tag, to generate pET21a-lipW. with 634bp fragment

APPENDIX II

CURRICULAR UPDATE

2013 – Gordon Research Conference- USA: SANTOS, D.F.K.; ISTVAN, P.; KRUGER, R.H. Antibiotic resistance genes from a soil metagenomic library and bioremediation applications.

2014 – VI Congresso Nacional de Química-Brasília. ISTVAN, P. Bio-prospection of lipase from soil metagenome and application on bio-remediation (Oral presentation).

2015 – VI Congresso Nacional de Biotecnologia-Brasília ISTVAN P, SANTOS, D.F.K.; KRUGER, R.H. Bioprospection of new lipolytic enzyme from Brazilian Cerrado soil metagenomic library and bioremediation applications

2015 – Article: New dioxygenase from metagenomic library from Brazilian soil: insights into antibiotic resistance and bioremediation. Biotechnology Letters (Dos Santos *et al.*, 2015)

2016 – Article: Functional metagenomics as a tool for identification of new antibiotic resistance genes (ARGs) from natural environments. Microbial. Ecology (Dos Santos *et al.*, 2016)

2017 Congresso brasileiro de Microbiologia ISTVAN, P., SILVA, T.F., KRÜGER, R.H. Molecular cloning and heterologous expression of a metagenomic esterase in *E. coli and Pichia Pastoris*

2018 – Technical visit at University Autonoma Barcelona, at Bioproces s e Biocatalisis Group: Molecular cloning and heterologous expression of a metagenomic esterase in *Komagataella phaffii P. pastoris*

APPENDIX III

Biotechnol Lett (2015) 37:1809–1817
DOI 10.1007/s10529-015-1861-x

ORIGINAL RESEARCH PAPER

New dioxygenase from metagenomic library from Brazilian soil: insights into antibiotic resistance and bioremediation

Débora Farage Knupp dos Santos · Paula Istvan · Eliane Ferreira Noronha · Betânia Ferraz Quirino · Ricardo Henrique Krüger

Received: 31 March 2015 / Accepted: 12 May 2015 / Published online: 21 May 2015

Abstract
Objectives Putative new dioxygenases were identified in a metagenomic β-lactam-resistance screening and, given their key role on aromatic metabolism, we raise the hypothesis that these enzymes maybe concomitantly related to antibiotic resistance and aromatic degradation.
Results ORFs of three putative dioxygenases were isolated from resistant metagenomic clones. One of them, CRB2(1), was subcloned into pET24a expression vector and subjected to downstream phenotypic and bioinformatics analyses that demonstrated the "dual effect" of our metagenomic dioxygenase, on antibiotic and aromatic resistance. Furthermore, initial characterization assays strongly suggests that CRB2(1) open-reading frame is an extradiol-dioxygenase, most probably a bicupin domain gentisate 1,2-dioxygenase. This observation is, to our knowledge, the first description of a metagenomic dioxygenase and its action on β-lactam resistance.
Conclusion Unraveling the diversity of antibiotic resistance elements on the environment could not only identify new genes and mechanisms in which bacteria can resist to antibiotics, but also contribute to biotechnology processes, such as in bioremediation.

Keywords Antibiotic resistance · Aromatic metabolism · Dioxygenase · Metagenome · Soil

D. F. K. dos Santos · P. Istvan · E. F. Noronha · R. H. Krüger
Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, Brazil
e-mail: deboratarage@yahoo.com.br

P. Istvan
e-mail: paula.istvan@gmail.com

E. F. Noronha
e-mail: enoronha@unb.br

B. F. Quirino
Embrapa-Agroenergia, Brasília, DF, Brazil
e-mail: betania.quirino@embrapa.br

B. F. Quirino
Genomic Sciences and Biotechnology Program, Universidade Católica de Brasília, Brasília, DF, Brazil

R. H. Krüger (✉)
Laboratório de Enzimologia, Instituto Central de Ciências Sul – Departamento de Biologia Celular, Universidade de Brasília – UnB, Brasília, DF CEP 700910-900, Brazil
e-mail: kruger@unb.br

Introduction

Resistance to antibiotics has been reported since the 1950s, shortly after their introduction for the treatment of infectious diseases (Davies 2007). The characterization of antibiotic resistance genes (ARGs) focuses primarily on clinical isolates; however, environmental habitats harbor an unappreciated and vast diversity of ARGs that are not yet well explored.

Therefore, it is urgent to identify and characterize ARGs from natural environments to better understand their diversity outside the clinical setting. We asked whether it was possible to identify new ARGs unrelated to clinical antibiotic resistance, and what their functions might be in natural environments.

In contrast to pathogenic microorganisms, most of which can be readily isolated and identified, only an estimated 0.1–1 % of soil microorganisms can be cultivated using traditional methods. This represents a major drawback for detecting ARGs in soil microorganisms, leaving culture-independent methods such as metagenomics as potential tools for accessing these microorganisms (Handelsman 2004). Here, we explore ARG diversity in a metagenomic library constructed from soil samples obtained in the Cerrado (de Castro et al. 2011), a savannah-like biome in the midwestern region of Brazil. In a functional screening performed with β-lactam antibiotics, 62 clones were isolated, and four were selected for sequencing reactions. Downstream analysis showed a large diversity of open reading frames (ORFs) inside the metagenomic inserts and assigned putative functions. In particular, genes encoding dioxygenase enzymes were frequently identified in the metagenomic clones.

Ring-cleaving dioxygenases play important roles in aromatic metabolism, both in eukaryotes and prokaryotes. These enzymes catalyze the incorporation of both atoms of O_2 into the aromatic ring, thereby cleaving it. Dioxygenases can be classified as intradiol dioxygenases, which catalyze ortho cleavage, and extradiol dioxygenases, which catalyze meta cleavage. Both classes of enzymes have distinct features and specificities (Arora et al. 2009), with extradiol dioxygenases generally being a more versatile group, cleaving a broader number of substrates (Vaillancourt et al. 2006).

Aromatic compounds are major pollutants discharged into the environment from different sources, including agricultural and industrial processes. These molecules are extremely stable in the environment, and existing physico-chemical procedures are often inefficient in their removal. Bioremediation is an efficient, cost-effective alternative that harness microbial degradation pathways to remove these persistent environmental pollutants via enzymatic catabolism (Arora et al. 2009).

This work tests the hypothesis that dioxygenases play a role in the resistance to β-lactamic antibiotics, and explores the dual effect of these enzymes in the resistance phenotype and aromatic metabolism, their primary function.

Materials and methods

Metagenomic libraries

The metagenomic libraries used in this work were previously constructed with Cerrado *stricto* sensu soil samples, as described by de Castro et al. (2011).

Resistance screening and subcloning

The functional screening of the metagenomic libraries, performed in lysogeny broth (LB) agar plates supplemented with nine β-lactamic antibiotics (16 μg amoxicillin ml^{-1}, 50 μg ampicillin ml^{-1}, 50 μg carbenicillin ml^{-1}, 16 μg cefamandole ml^{-1}, 20 μg cefoxitin ml^{-1}, 5 μg ceftazidime ml^{-1}, 50 μg cephalexin ml^{-1}, 50 μg penicillin G ml^{-1} and 12.5 μg piperacillin ml^{-1}), identified 62 resistant clones. Putative dioxygenase ORFs were subcloned into a pET24a vector, and transformed in *Escherichia coli* Tuner (DE3) cells.

Sequence analysis

Amino acid sequences were annotated using Protein BLAST (Altschul et al. 1990). Multiple alignments of the CRB2(1) amino acid sequence and related proteins were carried out using Clustal Omega (Sievers et al. 2011). Secondary structure was predicted using PSIPRED Protein Sequence Analysis Workbench (Buchan et al. 2013; Jones 1999). A phylogenetic tree was constructed with Mega 6.06 program (Tamura et al. 2013) using the neighbor-joining method and bootstrap analysis.

Expression and purification of putative metagenomic dioxygenase

To characterize the subclone, CRB2(1), that performed best in vector-based gene expression assays, cultures of it were grown to an OD_{600} of 0.6 and protein induction was carried out by overnight incubation with 1 % (w/v) lactose at 20 °C. The bacteria were lysed by sonication, and the protein of interest

Table 1 Protein BLAST annotation for putative dioxygenase open reading frames in three metagenomic subclones

Subclone	Selection antibiotic	Best hit	e value	Score	Similarity (%)	Coverage (%)	Conserved domains	Size	Estimated molecular mass (kDa)
AMX3(2)	Amoxicillin	Hypothetical protein (*Candidatus koribacter versatilis*)	2e-64	225	38	97	–	1077 bp 358 aa	39.8
AMX3(3)	Amoxicillin	Intradiol ring-cleavage dioxygenase (*Candidatus koribacter versatilis*)	2e-141	431	42	98	Peptidase_M14NE-CP-C_like superfamily	1674 bp 557 aa	61.9
CRB2(1)	Carbenicillin	Hypothetical protein (*Frankia* sp. Iso899)	2e-145	426	61	91	Cupin_2 superfamily	1068 bp 355 aa	39.5

was purified using the MagneHis Protein Purification System (Promega, Madison, WI, USA) under native conditions. SDS-PAGE analysis was performed in 4–13 % Tris–HCl gel and ran in denaturing running buffer pH 8.3. Samples were centrifuged, suspended in denaturing loading buffer 1× and heated for 5 min at 95 °C.

Phenotypic analysis and cell viability test

In order to evaluate resistance phenotypes to a β-lactamic antibiotic and phenol, CRB2(1) was grown in the presence of both compounds. Growth was monitored from the OD_{600} value. Cells carrying the expression vector containing subclone CRB2(1) were cultured at 28 °C with shaking (130 rpm). When cultures reached $OD_{600} = 0.1$, protein expression was induced by adding IPTG. After 1 h, 50 μg carbenicillin ml^{-1} and 0.1 % phenol were added, and OD_{600} was monitored from 1 to 96 h. This phenotypic assay was performed in quadruplicate. Cell viability was evaluated by adding 30 μl 2 mg MTT ml^{-1} to 200 μl of the 24 h cultures and incubating at 37 °C for 1 h.

Enzymatic assay

Gentisate 1,2-dioxygenase (GDO) activity was evaluated spectrophotometrically following the increase in absorption at 330 nm indicating maleylpyruvate formation. The 200 μl reaction mixture contained 50 μl purified putative dioxygenase (~35 μg), 130 μl phosphate buffered saline (0.1 M, pH 7.7) supplemented with 100 μM ferrous ammonium sulfate, and 2,5-dihydroxybenzoic acid (gentisic acid; 10 mM) as substrate. The samples were previously incubated at 40 °C for 1 h and then the absorbancy was measured every 20 min for 4 h. This assay was performed in triplicate. Blank samples consist of all the components but the purified protein, replaced by the elution buffer.

Results

Sequence analysis

Within the ORFs on the four selected metagenomics clones—AMX3, CFX12, CRB2 and PG17—we observed three putative dioxygenase genes. After subcloning processes, we constructed the respective subclones, named AMX3(2), AMX3(3) and CRB2(1) (Table 1). CRB2(1) subclone was selected for further analysis following some criteria—presence of conserved domains, better performance on IPTG induction assays and the presence of dioxygenase hits on the BLAST search.

Although the closest match from the BLAST search was a hypothetical protein, the CRB2(1) insert has a conserved bicupin domain, and GDO proteins were present within the first hundred Blastp hits. These findings suggest that CRB2(1) is an extradiol dioxygenase and could be further characterized as a novel GDO.

To identify the location of the cupin domains in the CRB2(1) protein, we used PSIPRED to predict its secondary structure (Fig. 1). The diverse cupin superfamily, which includes catalytic and non-catalytic proteins, is characterized by a six-stranded β-barrel fold. Catalytic members include dioxygenase enzymes. Two conserved motifs [$G(X)_5HXH(X)_{3-4}E(X)_6G$ and $G(X)_{5-7}PXG(X)_2H(X)_3$] comprise a cupin fold (Dunwell et al. 2001). Proteins with two cupin domains are called bicupins.

Because the secondary structure prediction identified two cupin domains, we aligned the CRB2(1) amino acid sequence with known bicupin dioxygenases (Fig. 2). Although all GDOs described to date contain two cupin domains (Adams et al. 2006; Chen et al. 2008; Hirano et al. 2007), the metal-binding sites are generally active in only one domain (Chen et al. 2008; Dunwell et al. 2004). Conserved metal-binding residues are present in both motifs of the cupin domain: two histidines and one glutamate in motif 1, and a histidine in motif 2. However, in GDOs the glutamate in motif 1 is typically replaced by alanine or another polar or hydrophobic residue (Fetzner 2012). In CRB2(1), an alanine-to-serine mutation occurred. The metal-binding site was conserved and is possibly active only in the *N*-terminal cupin domain of CRB2(1). Of the bicupin dioxygenases shown in Fig. 2, only the GDO of *Silicibacter pomeroyi* contains active metal-binding sites in both domains (Chen et al. 2008).

The phylogenetic tree distinguishes three groups of proteins: extradiol and intradiol dioxygenases and a monooxygenase (Fig. 3). CRB2(1) clusters with the extradiol dioxygenases and is placed along known cupin dioxygenases, including bicupin GDOs.

Functional analysis

To determine the resistance phenotype of CRB2(1), cells bearing the subclone were incubated with carbenicillin, the antibiotic initially used to select the CRB2 metagenomic clone, and phenol, a simple hydroxylated aromatic molecule with cytotoxic effects.

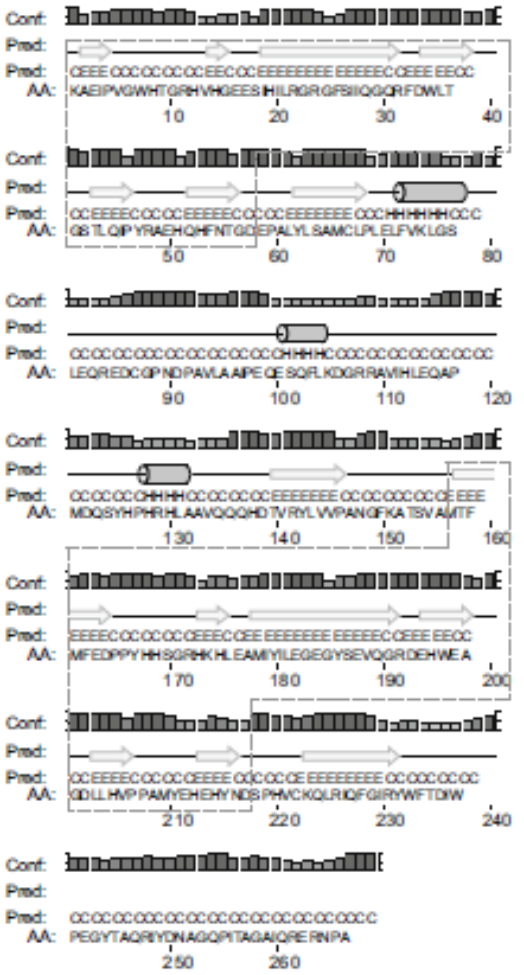


Fig. 1 Secondary structure prediction for the metagenomic clone CRB2(1) and characterization of cupin domains, represented by the *dashed polygons*. Each cupin domain contains six β-sheets, represented by *arrows*. The first and last two β-sheets contain motifs 1 and 2, respectively, for each cupin domain. This secondary structure was predicted with PSIPRED v3.3 and shows a bicupin domain-containing protein in the CRB2(1) amino acid sequence

Figure 4a demonstrates clear differences on cellular growth when CRB2(1) subclone and intact pET24a are compared. Then cell viability was evaluated after 24 h exposure to carbenicillin or phenol (Fig. 4b). Changes in the samples coloration shows the action of

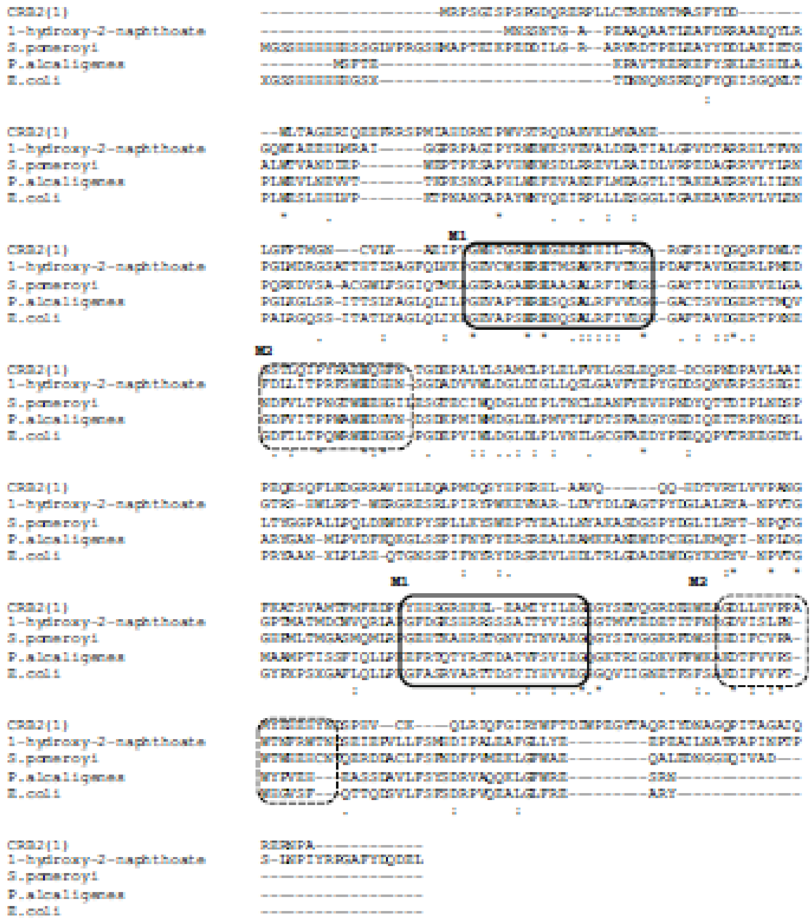


Fig. 2 Using the Clustal Omega multiple alignment tool, the CRB2(1) amino acid sequence was aligned with characterized gentisate 1,2-dioxygenases and a 1-hydroxy-2-naphtoate dioxygenase, all containing bicupin domains. Cupin motifs are represented by *rectangles*, divided by motif 1 (M1, represented by *bold squares*) and motif 2 (M2, represented by *dashed squares*). *Bold letters* represent conserved residues of the metal-binding motif of cupin domains (2-His-1Glu on motif 1 and 1-His on motif 2). In gentisate 1,2-dioxygenases the glutamate residue is typically replaced by an alanine; however, in CRB2(1) the glutamate residue was by a serine residue. Sequences downloaded from NCBI include 1-hydroxy-2-naphtoate from *Nocardioides sp.* K7 (BAA31235.2); gentisate 1,2-dioxygenase from *Silicibacter pomeroyi* (3BU7_A); gentisate 1,2-dioxygenase from *Escherichia coli* O57:H7 (2D40_A), and gentisate 1,2-dioxygenase from *Pseudomonas alcaligenes* (AAD49427.1)

living cells on MTT reagent and indicates cellular viability. These results demonstrate the dual effect of CRB2(1), conferring resistance to both an antibiotic and phenol.

Expression of the recombinant protein was induced with 1 % lactose at 20 °C for protein solubility, allowing for purification under native conditions. SDS-PAGE analysis of CRB2(1) suggests that it was a monomeric enzyme with an estimated molecular mass of 39.5 kDa, as predicted by sequence data (Fig. 5).

We then carried out an enzyme kinetics assay to evaluate the ability of the purified protein to cleave hydroxylated aromatic rings. Since we identified CRB2(1) as a potential GDO, gentisic acid was used as the substrate. Increased optical density at 330 nm in

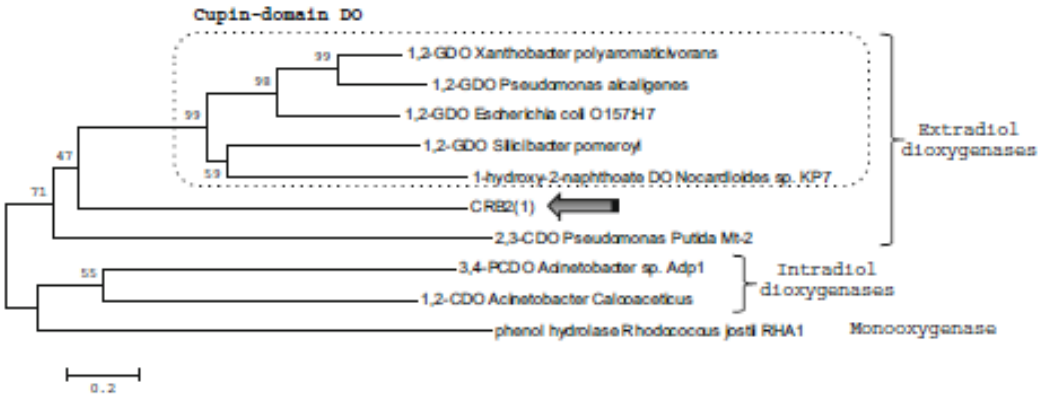


Fig. 3 Phylogenetic analysis of CRB2(1). Three distinct groups were identified and branched together: extradiol dioxygenases, intradiol dioxygenases, and a monooxygenase, represented by a phenol hydrolase. CRB2(1), indicated by the *arrow*, clusters in the extradiol dioxygenase group, along the subgroup of cupin domain dioxygenases (DO). The phylogenetic tree was constructed with the Mega 6.06 program, using the neighbor-joining method and bootstrap analysis (500 replicates). Sequences retrieved from NCBI include gentisate 1,2-dioxygenase (1,2-GDO) from *Xanthobacter polyaromaticivorans* (BAC98955.1); gentisate 1,2-dioxygenase (1,2-GDO) from *Pseudomonas alcaligenes* (AAD49427.1); gentisate 1,2-dioxygenase (1,2-GDO) from *Escherichia coli* O157:H7 (2D40_A); gentisate 1,2-dioxygenase (1,2-GDO) from *Silicibacter pomeroyi* (3BU7); 1-hydroxy-naphtoate-dioxygenase from *Nocardioides sp.* KP7 (BAA31235.2); catechol 2,3-dioxygenase (2,3-CDO) from *Pseudomonas putida* Mt-2 (1MPY_A); protocatechuate 3,4-dioxygenase (3,4-PCDO) from *Acinetobacter sp.* Adp1 (1EO2_A); catechol 1,2-dioxygenase (1,2-CDO) from *Acinetobacter calcoaceticus* (1DLM_B), and phenol hydrolase from *Rhodococcus jostii* RHA1 (YP_703833.1)

samples containing the recombinant protein indicated the accumulation of maleylpyruvate, formed by extradiol cleavage of gentisic acid (Fig. 6).

Discussion

Characterizing metagenomic DNA sequences presents some challenges. The most prominent is low sequence similarities with matches in existing databases; many are hypothetical or putative proteins, and conserved domains are generally absent. However, functional screening provides important clues about these sequences, with the substrate used in the initial assay being a strong indication for the classification of the ORF inside the insert. In this case, antibiotic resistance per se did not point to any particular ORF as the one responsible for the phenotype observed since the diversity of ARGs in the environment is not yet well understood. Our identification of putative dioxygenase genes in three metagenomic inserts suggested a possible role on the resistance phenotype. Subsequent tests not only confirm antibiotic resistance but also revealed the ability to cleave aromatic rings.

Several GDOs have been characterized, but none was identified in metagenomic sequences. Here we describe the initial characterization of a new dioxygenase isolated from an antibiotic-resistant metagenomic clone. Sequence and functional analyses indicate that CRB2(1) is likely a GDO with a conserved bicupin domain.

It is unclear why a dioxygenase was selected by a β-lactam-resistance screening of clones rather than well-characterized antibiotic resistance elements, such as β-lactamases. This finding supports the idea that ARG diversity, especially in the environment, is much greater than previously thought. In fact there is strong evidence that ARGs are not only widespread in the environment but predate the antibiotic era, as demonstrated by a metagenomic survey of ancient environmental DNA samples (D'Costa et al. 2011). This suggests that clinically important resistance elements were pre-existing genes that were selected by antibiotic use, with resistant strains likely adapting to increasing concentrations of these toxic compounds. Therefore, analysis of ARGs from natural environments may identify new resistance elements and help explain how these ARGs were transferred to

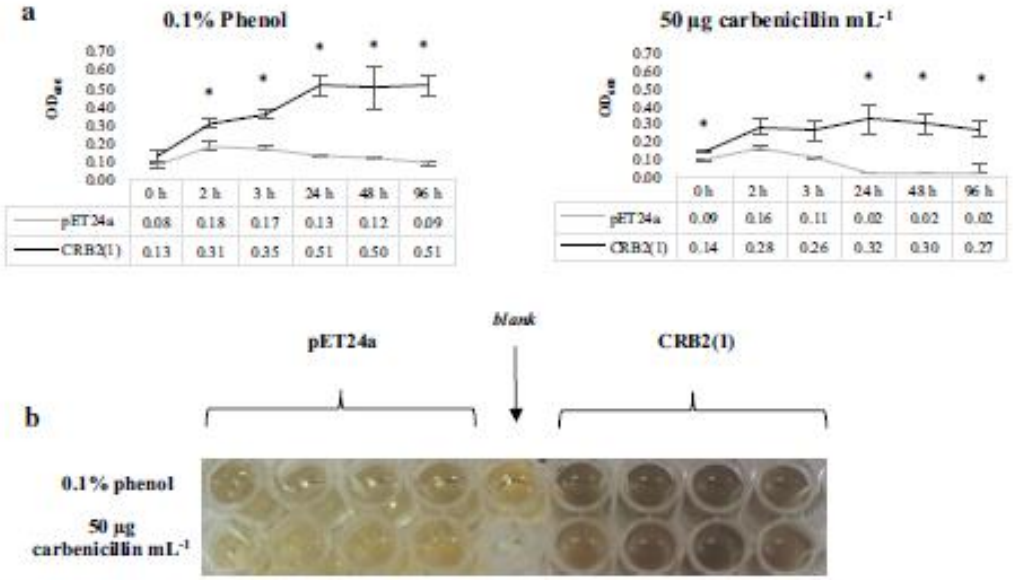


Fig. 4 **a** Cell density over time for cells containing pET24a vector containing CRB2(1) or empty vector on 50 µg carbenicillin ml^{-1} or 0.1 % phenol. At 0 h, protein expression was induced by adding IPTG, and 1 h later phenol or carbenicillin was added. Clear differences in turbidity and optical density values were observed by the 24 h time point. Results are expressed as mean values of quadruplicate samples. *$p < 0.01$. **b** Cell viability assay. Aliquots were obtained at the 24 h time point of cultures carrying the empty pET24a vector (*left*) and or vector containing the CRB2(1) subclone (*right*) after incubation with phenol or carbenicillin (each set of four wells represents assay replicates). MTT reduction (*dark color*) indicates viable cells, which is consistent with spectrophotometry results showing increased cell density for cells carrying the vector containing CRB2(1) and decreased cell density for cells carrying the empty vector after exposure to both compounds

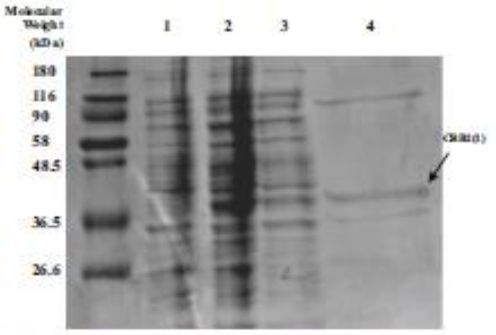


Fig. 5 SDS-PAGE analysis of the CRB2(1) subclone. *Lane 1*, cells carrying the empty pET24a (without CRB2(1) gene insert) after induction with 1 % lactose; *lane 2*, cells carrying pET24a vector with CRB2(1) after induction with 1 % lactose; *lane 3*, crude lysate supernatant from cells carrying vector with CRB2(1); *lane 4*, purified protein from CRB2(1) subclone. *Arrow* indicates the protein of interest. CRB2(1) insert band is absent in *lane 1*, showing the successful lactose induction and purification of CRB2(1) subclone

pathogenic microorganisms (e.g., via mobile elements such as transposons or integrons).

More attention should be turned to the origins of ARGs, especially those carried by non-cultivable microorganisms, which harbor a vast range of uncharacterized proteins, many with potential to cause antimicrobial resistance. For example, a recent functional metagenomic study on antimicrobial resistance in soil (Allen et al. 2015) revealed the role of a response-regulator gene on carbenicillin tolerance in *E. coli*, reinforcing the need to amplify our knowledge on antibiotic resistance elements, specially those from natural environments.

However, classification of a particular gene as an ARG can be difficult since its role in resistance may not be the primary function of the protein, as evidenced by our metagenomic clone.

This report is, to our knowledge, the first to show β-lactam-resistance activity of a dioxygenase. Although

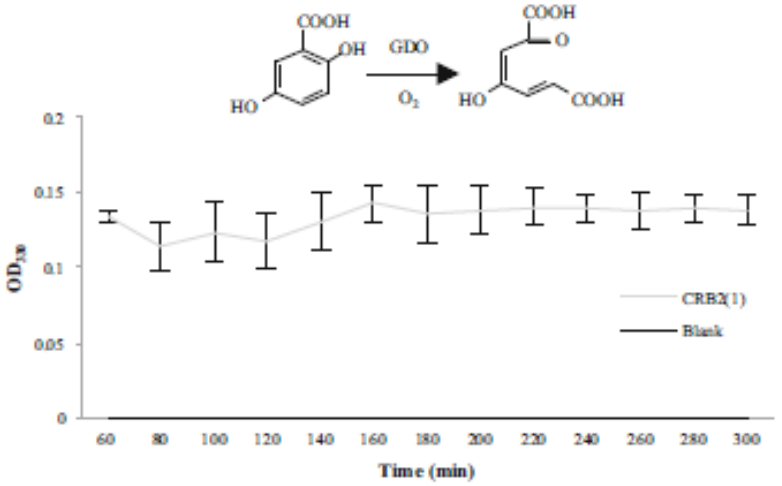


Fig. 6 Enzyme kinetics assay using purified CRB2(1) protein incubated with gentisic acid as substrate. Maleylpyruvate formation from the 2,3-cleavage of gentisic acid in the presence of oxygen was evaluated by spectrophotometry. The reads were performed after the 1 h incubation procedure. Blank samples lacked the purified protein. Results are expressed as mean ± SD of triplicate samples

the mechanism of antibiotic resistance is unclear, dioxygenases may act on the carbenicillin molecule by disrupting the aromatic ring, thereby inactivating its antibiotic activity. If so, dioxygenases represent another type of cleaving enzyme that confers β-lactam resistance, in addition to β-lactamases, which play an important role in clinical infections.

The ability of this protein to degrade aromatic rings, as evidenced by phenol resistance and the kinetic assay, is another important finding of this work. Antimicrobial resistance genes are frequently found in association with other genes that confer similar phenotypes, such as resistance to heavy metals or quaternary ammonium. Aromatic degradation may play an analogous role, selecting other genes to construct xenobiotic resistance gene cassettes. The phenol resistance of this metagenomic clone also suggests its possible use in bioremediation. Aromatic compounds are important environmental pollutants, and degradation by microbial enzymes can facilitate their removal from contaminated areas. Thus, accessing microbial diversity, especially that of non-cultivable and consequently unknown microorganisms, could increase the range of microbial enzymes used in this field.

Acknowledgments The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) for the financial support.

References

Adams MA, Singh VK, Keller BO, Jia Z (2006) Structural and biochemical characterization of gentisate 1,2-dioxygenase from *Escherichia coli* O157:H7. Mol Microbiol 61:1469–1484

Allen HK, An R, Handelsman J, Moe LA (2015) A response regulator from a soil metagenome enhances resistance to the β-lactam antibiotic carbenicillin in *Escherichia coli*. PLoS One 10(3):e0120094

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arora PK, Kumar M, Chauhan A, Raghava GP, Jain RK (2009) OxDBase: a database of oxygenases involved in biodegradation. BMC Res Notes 2:67

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. Nucl Acid Res 41:W349–W357

Chen J, Li W, Wang M, Zhu G et al (2008) Crystal structure and mutagenic analysis of GDOsp, a gentisate 1,2-dioxygenase from *Silicibacter pomeroyi*. Protein Sci 17:1362–1373

Davies J (2007) Microbes have the last word. A drastic re-evaluation of antimicrobial treatment is needed to overcome the threat of antibiotic-resistant bacteria. EMBO Rep 8:616–621

D'Costa VM, King CE, Kalan L, Morar M et al (2011) Antibiotic resistance is ancient. Nature 477:457–461

de Castro AP, Quirino BF, Allen H, Williamson LL, Handelsman J, Kruger RH (2011) Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content. Biotechnol Lett 33:2169–2175

Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW (2001) Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci 26:740–746

Dunwell JM, Purvis A, Khuri S (2004) Cupins: the most functionally diverse protein superfamily? Phytochemistry 65:7–17

APPENDIX IV

Biotechnol Lett (2015) 37:1809–1817
DOI 10.1007/s10529-015-1861-x

ORIGINAL RESEARCH PAPER

New dioxygenase from metagenomic library from Brazilian soil: insights into antibiotic resistance and bioremediation

Débora Farage Knupp dos Santos · Paula Istvan · Eliane Ferreira Noronha · Betânia Ferraz Quirino · Ricardo Henrique Krüger

Received: 31 March 2015 / Accepted: 12 May 2015 / Published online: 21 May 2015

Abstract
Objectives Putative new dioxygenases were identified in a metagenomic β-lactam-resistance screening and, given their key role on aromatic metabolism, we raise the hypothesis that these enzymes maybe concomitantly related to antibiotic resistance and aromatic degradation.
Results ORFs of three putative dioxygenases were isolated from resistant metagenomic clones. One of them, CRB2(1), was subcloned into pET24a expression vector and subjected to downstream phenotypic and bioinformatics analyses that demonstrated the "dual effect" of our metagenomic dioxygenase, on antibiotic and aromatic resistance. Furthermore, initial characterization assays strongly suggests that CRB2(1) open-reading frame is an extradiol-dioxygenase, most probably a bicupin domain gentisate 1,2-dioxygenase. This observation is, to our knowledge, the first description of a metagenomic dioxygenase and its action on β-lactam resistance.
Conclusion Unraveling the diversity of antibiotic resistance elements on the environment could not only identify new genes and mechanisms in which bacteria can resist to antibiotics, but also contribute to biotechnology processes, such as in bioremediation.

Keywords Antibiotic resistance · Aromatic metabolism · Dioxygenase · Metagenome · Soil

D. F. K. dos Santos · P. Istvan · E. F. Noronha · R. H. Krüger
Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, Brazil
e-mail: deboratarage@yahoo.com.br

P. Istvan
e-mail: paula.istvan@gmail.com

E. F. Noronha
e-mail: enoronha@unb.br

B. F. Quirino
Embrapa-Agroenergia, Brasília, DF, Brazil
e-mail: betania.quirino@embrapa.br

B. F. Quirino
Genomic Sciences and Biotechnology Program, Universidade Católica de Brasília, Brasília, DF, Brazil

R. H. Krüger (✉)
Laboratório de Enzimologia, Instituto Central de Ciências Sul – Departamento de Biologia Celular, Universidade de Brasília – UnB, Brasília, DF CEP 700910-900, Brazil
e-mail: kruger@unb.br

Introduction

Resistance to antibiotics has been reported since the 1950s, shortly after their introduction for the treatment of infectious diseases (Davies 2007). The characterization of antibiotic resistance genes (ARGs) focuses primarily on clinical isolates; however, environmental habitats harbor an unappreciated and vast diversity of ARGs that are not yet well explored.

Therefore, it is urgent to identify and characterize ARGs from natural environments to better understand their diversity outside the clinical setting. We asked whether it was possible to identify new ARGs unrelated to clinical antibiotic resistance, and what their functions might be in natural environments.

In contrast to pathogenic microorganisms, most of which can be readily isolated and identified, only an estimated 0.1–1 % of soil microorganisms can be cultivated using traditional methods. This represents a major drawback for detecting ARGs in soil microorganisms, leaving culture-independent methods such as metagenomics as potential tools for accessing these microorganisms (Handelsman 2004). Here, we explore ARG diversity in a metagenomic library constructed from soil samples obtained in the Cerrado (de Castro et al. 2011), a savannah-like biome in the midwestern region of Brazil. In a functional screening performed with β-lactam antibiotics, 62 clones were isolated, and four were selected for sequencing reactions. Downstream analysis showed a large diversity of open reading frames (ORFs) inside the metagenomic inserts and assigned putative functions. In particular, genes encoding dioxygenase enzymes were frequently identified in the metagenomic clones.

Ring-cleaving dioxygenases play important roles in aromatic metabolism, both in eukaryotes and prokaryotes. These enzymes catalyze the incorporation of both atoms of O_2 into the aromatic ring, thereby cleaving it. Dioxygenases can be classified as intradiol dioxygenases, which catalyze ortho cleavage, and extradiol dioxygenases, which catalyze meta cleavage. Both classes of enzymes have distinct features and specificities (Arora et al. 2009), with extradiol dioxygenases generally being a more versatile group, cleaving a broader number of substrates (Vaillancourt et al. 2006).

Aromatic compounds are major pollutants discharged into the environment from different sources, including agricultural and industrial processes. These molecules are extremely stable in the environment, and existing physico-chemical procedures are often inefficient in their removal. Bioremediation is an efficient, cost-effective alternative that harness microbial degradation pathways to remove these persistent environmental pollutants via enzymatic catabolism (Arora et al. 2009).

This work tests the hypothesis that dioxygenases play a role in the resistance to β-lactamic antibiotics, and explores the dual effect of these enzymes in the resistance phenotype and aromatic metabolism, their primary function.

Materials and methods

Metagenomic libraries

The metagenomic libraries used in this work were previously constructed with Cerrado *stricto* sensu soil samples, as described by de Castro et al. (2011).

Resistance screening and subcloning

The functional screening of the metagenomic libraries, performed in lysogeny broth (LB) agar plates supplemented with nine β-lactamic antibiotics (16 μg amoxicillin ml^{-1}, 50 μg ampicillin ml^{-1}, 50 μg carbenicillin ml^{-1}, 16 μg cefamandole ml^{-1}, 20 μg cefoxitin ml^{-1}, 5 μg ceftazidime ml^{-1}, 50 μg cephalexin ml^{-1}, 50 μg penicillin G ml^{-1} and 12.5 μg piperacillin ml^{-1}), identified 62 resistant clones. Putative dioxygenase ORFs were subcloned into a pET24a vector, and transformed in *Escherichia coli* Tuner (DE3) cells.

Sequence analysis

Amino acid sequences were annotated using Protein BLAST (Altschul et al. 1990). Multiple alignments of the CRB2(1) amino acid sequence and related proteins were carried out using Clustal Omega (Sievers et al. 2011). Secondary structure was predicted using PSIPRED Protein Sequence Analysis Workbench (Buchan et al. 2013; Jones 1999). A phylogenetic tree was constructed with Mega 6.06 program (Tamura et al. 2013) using the neighbor-joining method and bootstrap analysis.

Expression and purification of putative metagenomic dioxygenase

To characterize the subclone, CRB2(1), that performed best in vector-based gene expression assays, cultures of it were grown to an OD_{600} of 0.6 and protein induction was carried out by overnight incubation with 1 % (w/v) lactose at 20 °C. The bacteria were lysed by sonication, and the protein of interest

Table 1 Protein BLAST annotation for putative dioxygenase open reading frames in three metagenomic subclones

Subclone	Selection antibiotic	Best hit	e value	Score	Similarity (%)	Coverage (%)	Conserved domains	Size	Estimated molecular mass (kDa)
AMX3(2)	Amoxicillin	Hypothetical protein (*Candidatus koribacter versatilis*)	2e-64	225	38	97	–	1077 bp 358 aa	39.8
AMX3(3)	Amoxicillin	Intradiol ring-cleavage dioxygenase (*Candidatus koribacter versatilis*)	2e-141	431	42	98	Peptidase_M14NE-CP-C_like superfamily	1674 bp 557 aa	61.9
CRB2(1)	Carbenicillin	Hypothetical protein (*Frankia* sp. Iso899)	2e-145	426	61	91	Cupin_2 superfamily	1068 bp 355 aa	39.5

was purified using the MagneHis Protein Purification System (Promega, Madison, WI, USA) under native conditions. SDS-PAGE analysis was performed in 4–13 % Tris–HCl gel and ran in denaturing running buffer pH 8.3. Samples were centrifuged, suspended in denaturing loading buffer 1× and heated for 5 min at 95 °C.

Phenotypic analysis and cell viability test

In order to evaluate resistance phenotypes to a β-lactamic antibiotic and phenol, CRB2(1) was grown in the presence of both compounds. Growth was monitored from the OD_{600} value. Cells carrying the expression vector containing subclone CRB2(1) were cultured at 28 °C with shaking (130 rpm). When cultures reached $OD_{600} = 0.1$, protein expression was induced by adding IPTG. After 1 h, 50 μg carbenicillin ml^{-1} and 0.1 % phenol were added, and OD_{600} was monitored from 1 to 96 h. This phenotypic assay was performed in quadruplicate. Cell viability was evaluated by adding 30 μl 2 mg MTT ml^{-1} to 200 μl of the 24 h cultures and incubating at 37 °C for 1 h.

Enzymatic assay

Gentisate 1,2-dioxygenase (GDO) activity was evaluated spectrophotometrically following the increase in absorption at 330 nm indicating maleylpyruvate formation. The 200 μl reaction mixture contained 50 μl purified putative dioxygenase (~35 μg), 130 μl phosphate buffered saline (0.1 M, pH 7.7) supplemented with 100 μM ferrous ammonium sulfate, and 2,5-dihydroxybenzoic acid (gentisic acid; 10 mM) as substrate. The samples were previously incubated at 40 °C for 1 h and then the absorbancy was measured every 20 min for 4 h. This assay was performed in triplicate. Blank samples consist of all the components but the purified protein, replaced by the elution buffer.

Results

Sequence analysis

Within the ORFs on the four selected metagenomics clones—AMX3, CFX12, CRB2 and PG17—we observed three putative dioxygenase genes. After subcloning processes, we constructed the respective subclones, named AMX3(2), AMX3(3) and CRB2(1) (Table 1). CRB2(1) subclone was selected for further analysis following some criteria—presence of conserved domains, better performance on IPTG induction assays and the presence of dioxygenase hits on the BLAST search.

Although the closest match from the BLAST search was a hypothetical protein, the CRB2(1) insert has a conserved bicupin domain, and GDO proteins were present within the first hundred Blastp hits. These findings suggest that CRB2(1) is an extradiol dioxygenase and could be further characterized as a novel GDO.

To identify the location of the cupin domains in the CRB2(1) protein, we used PSIPRED to predict its secondary structure (Fig. 1). The diverse cupin superfamily, which includes catalytic and non-catalytic proteins, is characterized by a six-stranded β-barrel fold. Catalytic members include dioxygenase enzymes. Two conserved motifs [$G(X)_5HXH(X)_{3-4}E(X)_6G$ and $G(X)_{5-7}PXG(X)_2H(X)_3$] comprise a cupin fold (Dunwell et al. 2001). Proteins with two cupin domains are called bicupins.

Because the secondary structure prediction identified two cupin domains, we aligned the CRB2(1) amino acid sequence with known bicupin dioxygenases (Fig. 2). Although all GDOs described to date contain two cupin domains (Adams et al. 2006; Chen et al. 2008; Hirano et al. 2007), the metal-binding sites are generally active in only one domain (Chen et al. 2008; Dunwell et al. 2004). Conserved metal-binding residues are present in both motifs of the cupin domain: two histidines and one glutamate in motif 1, and a histidine in motif 2. However, in GDOs the glutamate in motif 1 is typically replaced by alanine or another polar or hydrophobic residue (Fetzner 2012). In CRB2(1), an alanine-to-serine mutation occurred. The metal-binding site was conserved and is possibly active only in the *N*-terminal cupin domain of CRB2(1). Of the bicupin dioxygenases shown in Fig. 2, only the GDO of *Silicibacter pomeroyi* contains active metal-binding sites in both domains (Chen et al. 2008).

The phylogenetic tree distinguishes three groups of proteins: extradiol and intradiol dioxygenases and a monooxygenase (Fig. 3). CRB2(1) clusters with the extradiol dioxygenases and is placed along known cupin dioxygenases, including bicupin GDOs.

Functional analysis

To determine the resistance phenotype of CRB2(1), cells bearing the subclone were incubated with carbenicillin, the antibiotic initially used to select the CRB2 metagenomic clone, and phenol, a simple

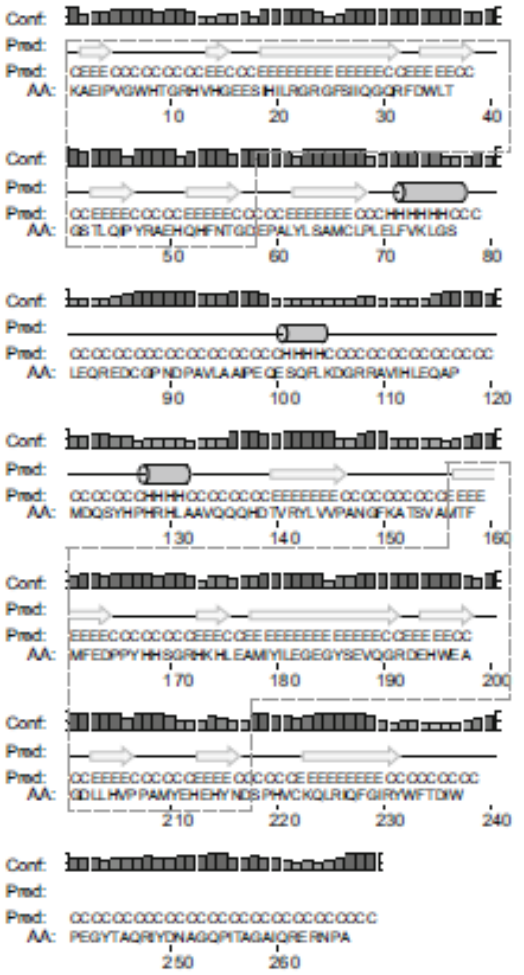


Fig. 1 Secondary structure prediction for the metagenomic clone CRB2(1) and characterization of cupin domains, represented by the *dashed polygons*. Each cupin domain contains six β-sheets, represented by *arrows*. The first and last two β-sheets contain motifs 1 and 2, respectively, for each cupin domain. This secondary structure was predicted with PSIPRED v3.3 and shows a bicupin domain-containing protein in the CRB2(1) amino acid sequence

hydroxylated aromatic molecule with cytotoxic effects.

Figure 4a demonstrates clear differences on cellular growth when CRB2(1) subclone and intact pET24a are compared. Then cell viability was evaluated after 24 h exposure to carbenicillin or phenol (Fig. 4b). Changes in the samples coloration shows the action of

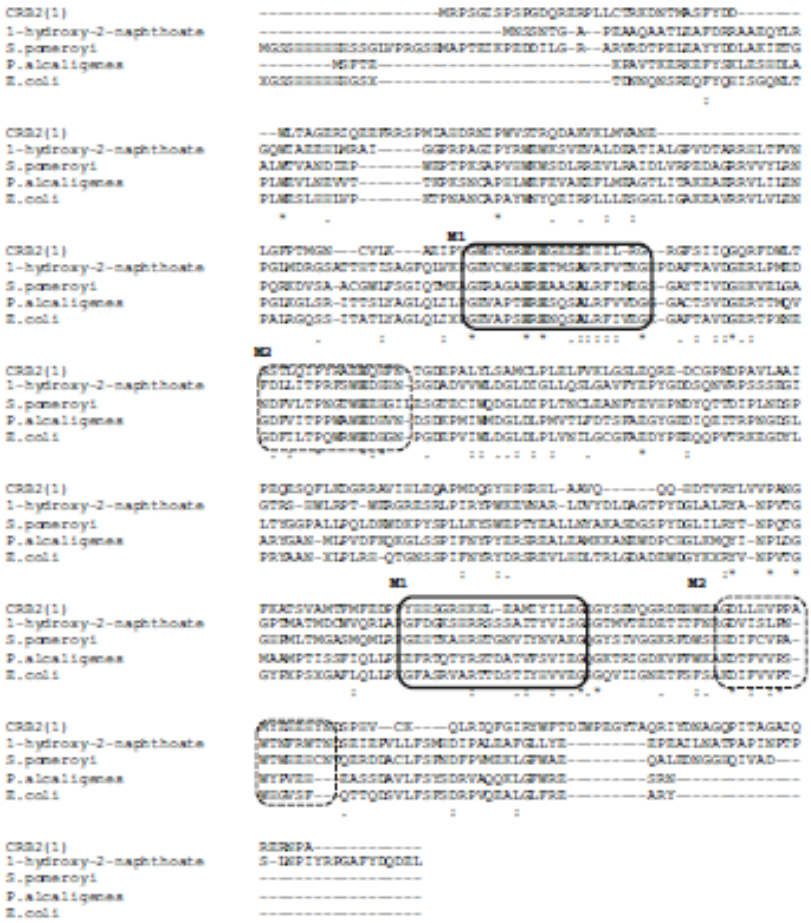


Fig. 2 Using the Clustal Omega multiple alignment tool, the CRB2(1) amino acid sequence was aligned with characterized gentisate 1,2-dioxygenases and a 1-hydroxy-2-naphtoate dioxygenase, all containing bicupin domains. Cupin motifs are represented by *rectangles*, divided by motif 1 (M1, represented by *bold squares*) and motif 2 (M2, represented by *dashed squares*). *Bold letters* represent conserved residues of the metal-binding motif of cupin domains (2-His-1Glu on motif 1 and 1-His on motif 2). In gentisate 1,2-dioxygenases the glutamate residue is typically replaced by an alanine; however, in CRB2(1) the glutamate residue was bya serine residue. Sequences downloaded from NCBI include 1-hydroxy-2-naphtoate from *Novosphingobium sp.* K7 (BAA31235.2); gentisate 1,2-dioxygenase from *Silicibacter pomeroyi* (3BU7_A); gentisate 1,2-dioxygenase from *Escherichia coli* O57:H7 (2D40_A), and gentisate 1,2-dioxygenase from *Pseudomonas alcaligenes* (AAD49427.1)

living cells on MTT reagent and indicates cellular viability. These results demonstrate the dual effect of CRB2(1), conferring resistance to both an antibiotic and phenol.

Expression of the recombinant protein was induced with 1 % lactose at 20 °C for protein solubility, allowing for purification under native conditions. SDS-PAGE analysis of CRB2(1) suggests that it was a monomeric enzyme with an estimated molecular mass of 39.5 kDa, as predicted by sequence data (Fig. 5).

We then carried out an enzyme kinetics assay to evaluate the ability of the purified protein to cleave hydroxylated aromatic rings. Since we identified CRB2(1) as a potential GDO, gentisic acid was used as the substrate. Increased optical density at 330 nm in

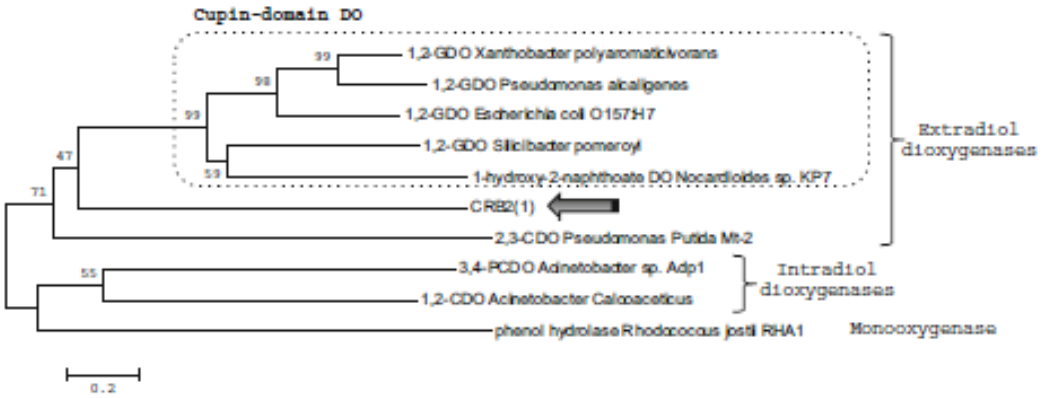


Fig. 3 Phylogenetic analysis of CRB2(1). Three distinct groups were identified and branched together: extradiol dioxygenases, intradiol dioxygenases, and a monooxygenase, represented by a phenol hydrolase. CRB2(1), indicated by the *arrow*, clusters in the extradiol dioxygenase group, along the subgroup of cupin domain dioxygenases (DO). The phylogenetic tree was constructed with the Mega 6.06 program, using the neighbor-joining method and bootstrap analysis (500 replicates). Sequences retrieved from NCBI include gentisate 1,2-dioxygenase (1,2-GDO) from *Xanthobacter polyaromaticivorans* (BAC98955.1); gentisate 1,2-dioxygenase (1,2-GDO) from *Pseudomonas alcaligenes* (AAD49427.1); gentisate 1,2-dioxygenase (1,2-GDO) from *Escherichia coli* O157:H7 (2D40_A); gentisate 1,2-dioxygenase (1,2-GDO) from *Silicibacter pomeroyi* (3BU7); 1-hydroxy-naphtoate-dioxygenase from *Nocardioides sp.* KP7 (BAA31235.2); catechol 2,3-dioxygenase (2,3-CDO) from *Pseudomonas putida* Mt-2 (1MPY_A); protocatechuate 3,4-dioxygenase (3,4-PCDO) from *Acinetobacter sp.* Adp1 (1EO2_A); catechol 1,2-dioxygenase (1,2-CDO) from *Acinetobacter calcoaceticus* (1DLM_B), and phenol hydrolase from *Rhodococcus jostii* RHA1 (YP_703833.1)

samples containing the recombinant protein indicated the accumulation of maleylpyruvate, formed by extradiol cleavage of gentisic acid (Fig. 6).

Discussion

Characterizing metagenomic DNA sequences presents some challenges. The most prominent is low sequence similarities with matches in existing databases; many are hypothetical or putative proteins, and conserved domains are generally absent. However, functional screening provides important clues about these sequences, with the substrate used in the initial assay being a strong indication for the classification of the ORF inside the insert. In this case, antibiotic resistance per se did not point to any particular ORF as the one responsible for the phenotype observed since the diversity of ARGs in the environment is not yet well understood. Our identification of putative dioxygenase genes in three metagenomic inserts suggested a possible role on the resistance phenotype. Subsequent tests not only confirm antibiotic resistance but also revealed the ability to cleave aromatic rings.

Several GDOs have been characterized, but none was identified in metagenomic sequences. Here we describe the initial characterization of a new dioxygenase isolated from an antibiotic-resistant metagenomic clone. Sequence and functional analyses indicate that CRB2(1) is likely a GDO with a conserved bicupin domain.

It is unclear why a dioxygenase was selected by a β-lactam-resistance screening of clones rather than well-characterized antibiotic resistance elements, such as β-lactamases. This finding supports the idea that ARG diversity, especially in the environment, is much greater than previously thought. In fact there is strong evidence that ARGs are not only widespread in the environment but predate the antibiotic era, as demonstrated by a metagenomic survey of ancient environmental DNA samples (D'Costa et al. 2011). This suggests that clinically important resistance elements were pre-existing genes that were selected by antibiotic use, with resistant strains likely adapting to increasing concentrations of these toxic compounds. Therefore, analysis of ARGs from natural environments may identify new resistance elements and help explain how these ARGs were transferred to

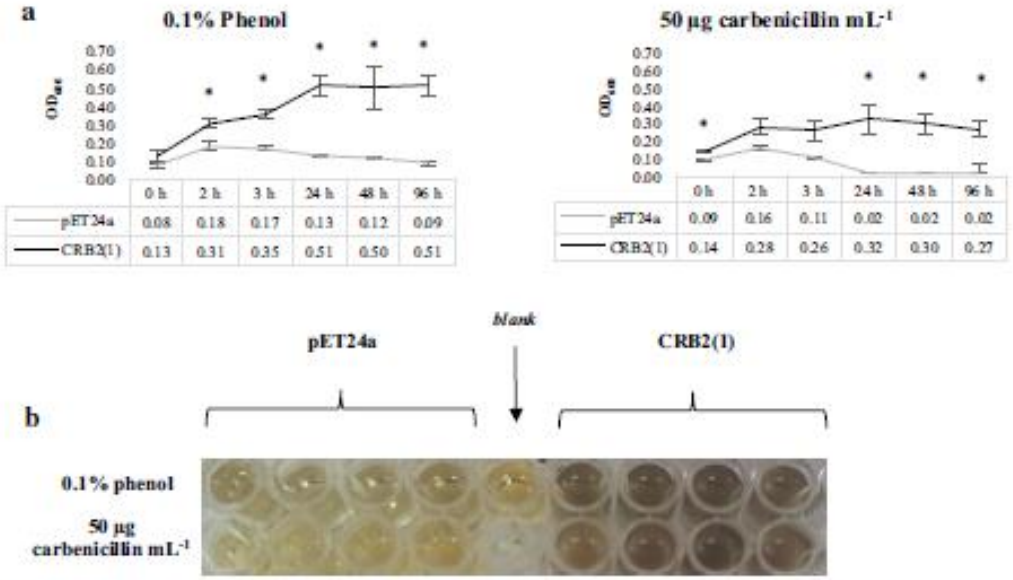


Fig. 4 **a** Cell density over time for cells containing pET24a vector containing CRB2(1) or empty vector on 50 µg carbenicillin ml^{-1} or 0.1 % phenol. At 0 h, protein expression was induced by adding IPTG, and 1 h later phenol or carbenicillin was added. Clear differences in turbidity and optical density values were observed by the 24 h time point. Results are expressed as mean values of quadruplicate samples. *$p < 0.01$. **b** Cell viability assay. Aliquots were obtained at the 24 h time point of cultures carrying the empty pET24a vector (*left*) and or vector containing the CRB2(1) subclone (*right*) after incubation with phenol or carbenicillin (each set of four wells represents assay replicates). MTT reduction (*dark color*) indicates viable cells, which is consistent with spectrophotometry results showing increased cell density for cells carrying the vector containing CRB2(1) and decreased cell density for cells carrying the empty vector after exposure to both compounds

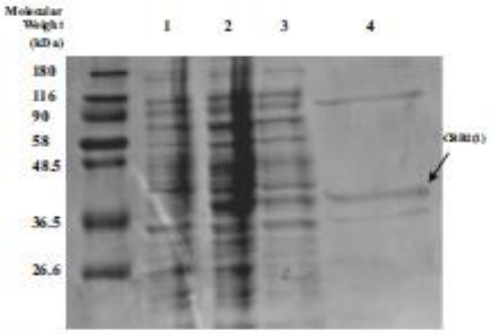


Fig. 5 SDS-PAGE analysis of the CRB2(1) subclone. *Lane 1*, cells carrying the empty pET24a (without CRB2(1) gene insert) after induction with 1 % lactose; *lane 2*, cells carrying pET24a vector with CRB2(1) after induction with 1 % lactose; *lane 3*, crude lysate supernatant from cells carrying vector with CRB2(1); *lane 4*, purified protein from CRB2(1) subclone. *Arrow* indicates the protein of interest. CRB2(1) insert band is absent in *lane 1*, showing the successful lactose induction and purification of CRB2(1) subclone

pathogenic microorganisms (e.g., via mobile elements such as transposons or integrons).

More attention should be turned to the origins of ARGs, especially those carried by non-cultivable microorganisms, which harbor a vast range of uncharacterized proteins, many with potential to cause antimicrobial resistance. For example, a recent functional metagenomic study on antimicrobial resistance in soil (Allen et al. 2015) revealed the role of a response-regulator gene on carbenicillin tolerance in *E. coli*, reinforcing the need to amplify our knowledge on antibiotic resistance elements, specially those from natural environments.

However, classification of a particular gene as an ARG can be difficult since its role in resistance may not be the primary function of the protein, as evidenced by our metagenomic clone.

This report is, to our knowledge, the first to show β-lactam-resistance activity of a dioxygenase. Although

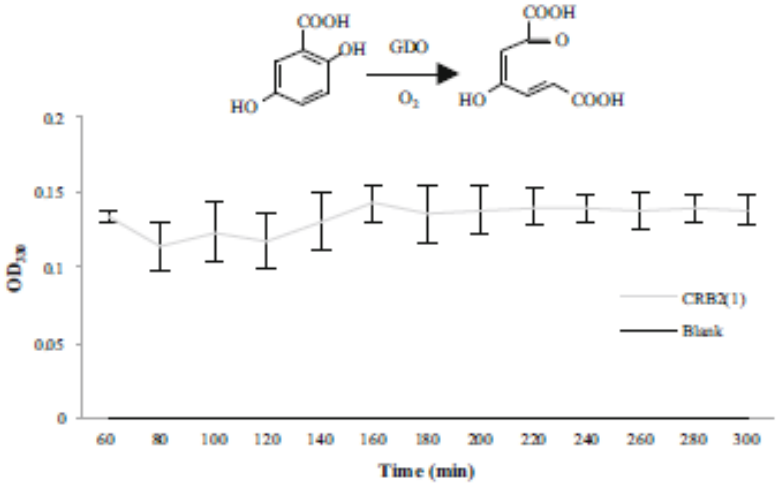


Fig. 6 Enzyme kinetics assay using purified CRB2(1) protein incubated with gentisic acid as substrate. Maleylpyruvate formation from the 2,3-cleavage of gentisic acid in the presence of oxygen was evaluated by spectrophotometry. The reads were performed after the 1 h incubation procedure. Blank samples lacked the purified protein. Results are expressed as mean ± SD of triplicate samples

the mechanism of antibiotic resistance is unclear, dioxygenases may act on the carbenicillin molecule by disrupting the aromatic ring, thereby inactivating its antibiotic activity. If so, dioxygenases represent another type of cleaving enzyme that confers β-lactam resistance, in addition to β-lactamases, which play an important role in clinical infections.

The ability of this protein to degrade aromatic rings, as evidenced by phenol resistance and the kinetic assay, is another important finding of this work. Antimicrobial resistance genes are frequently found in association with other genes that confer similar phenotypes, such as resistance to heavy metals or quaternary ammonium. Aromatic degradation may play an analogous role, selecting other genes to construct xenobiotic resistance gene cassettes. The phenol resistance of this metagenomic clone also suggests its possible use in bioremediation. Aromatic compounds are important environmental pollutants, and degradation by microbial enzymes can facilitate their removal from contaminated areas. Thus, accessing microbial diversity, especially that of non-cultivable and consequently unknown microorganisms, could increase the range of microbial enzymes used in this field.

Acknowledgments The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF) for the financial support.

References

Adams MA, Singh VK, Keller BO, Jia Z (2006) Structural and biochemical characterization of gentisate 1,2-dioxygenase from *Escherichia coli* O157:H7. Mol Microbiol 61:1469–1484

Allen HK, An R, Handelsman J, Moe LA (2015) A response regulator from a soil metagenome enhances resistance to the β-lactam antibiotic carbenicillin in *Escherichia coli*. PLoS One 10(3):e0120094

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arora PK, Kumar M, Chauhan A, Raghava GP, Jain RK (2009) OxDBase: a database of oxygenases involved in biodegradation. BMC Res Notes 2:67

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. Nucl Acid Res 41:W349–W357

Chen J, Li W, Wang M, Zhu G et al (2008) Crystal structure and mutagenic analysis of GDOsp, a gentisate 1,2-dioxygenase from *Silicibacter pomeroyi*. Protein Sci 17:1362–1373

Davies J (2007) Microbes have the last word. A drastic re-evaluation of antimicrobial treatment is needed to overcome the threat of antibiotic-resistant bacteria. EMBO Rep 8:616–621

D'Costa VM, King CE, Kalan L, Morar M et al (2011) Antibiotic resistance is ancient. Nature 477:457–461

de Castro AP, Quirino BF, Allen H, Williamson LL, Handelsman J, Kruger RH (2011) Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content. Biotechnol Lett 33:2169–2175

Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW (2001) Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci 26:740–746

Dunwell JM, Purvis A, Khuri S (2004) Cupins: the most functionally diverse protein superfamily? Phytochemistry 65:7–17

Fetzner S (2012) Ring-cleaving dioxygenases with a cupin fold. Appl Environ Microbiol 78:2505–2514

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Hirano S, Morikawa M, Takano K, Imanaka T, Kanaya S (2007) Gentisate 1,2-dioxygenase from *Xanthobacter polyaromaticivorans* 127W. Biosci Biotechnol Biochem 71:192–199

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195–202

Sievers F et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Vaillancourt FH, Bolin JT, Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol 41:241–267

APPENDIX V

Microb Ecol
DOI 10.1007/s00248-016-0866-x

MINIREVIEWS

Functional Metagenomics as a Tool for Identification of New Antibiotic Resistance Genes from Natural Environments

Débora Farage Knupp dos Santos[1] · Paula Istvan[1] · Betania Ferraz Quirino[2,3] · Ricardo Henrique Kruger[1]

Received: 18 January 2016 / Accepted: 19 September 2016

Abstract Antibiotic resistance has become a major concern for human and animal health, as therapeutic alternatives to treat multidrug-resistant microorganisms are rapidly dwindling. The problem is compounded by low investment in antibiotic research and lack of new effective antimicrobial drugs on the market. Exploring environmental antibiotic resistance genes (ARGs) will help us to better understand bacterial resistance mechanisms, which may be the key to identifying new drug targets. Because most environment-associated microorganisms are not yet cultivable, culture-independent techniques are essential to determine which organisms are present in a given environmental sample and allow the assessment and utilization of the genetic wealth they represent. Metagenomics represents a powerful tool to achieve these goals using sequence-based and functional-based approaches. Functional metagenomic approaches are particularly well suited to the identification new ARGs from natural environments because, unlike sequence-based approaches, they do not require previous knowledge of these genes. This review discusses functional metagenomics-based ARG research and describes new possibilities for surveying the resistome in environmental samples.

Keywords Antibiotic resistance genes · Environment · Functional metagenomics · Resistome

✉ Ricardo Henrique Kruger
kruger@unb.br

[1] Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, Brazil

[2] Embrapa-Agroenergia, Brasília, DF, Brazil

[3] Universidade Católica de Brasília, Genomic Sciences and Biotechnology Program, Brasília, DF, Brazil

Published online: 05 October 2016

Introduction

The release of new antimicrobials by the pharmaceutical industry has decreased over the years compared to that of other drug classes such as anticancer drugs. Although the Food and Drug Administration (FDA) approved 40 new anticancer agents in the last 6 years, only 8 new antibiotic formulations were approved during that time (available at www.fda.gov) (Fig. 1). Moreover, none of the new molecules belong to novel classes of antibiotics but are derivatives of already available chemical scaffolds. In fact, no real innovation has occurred in the antibiotic drug industry since 2009, when telavancin (Vibativ®; Theravance, Inc.) was released onto the market. This drug was the first lipoglycopeptide antibiotic approved by the FDA for treating complicated skin infections and ventilator-associated pneumonia (available at www.fda.gov). Although progress has been made in antibiotic discovery, such as the platform presented by Seiple, I.B. and collaborators [1], this scenario remains of great concern to public health, as multidrug-resistant pathogenic microorganisms are becoming widespread, while drugs able to control a possible epidemic are lacking. Thus, new molecules to combat resistant bacteria are urgently needed.

Although antibiotic resistance in pathogenic and clinically important bacteria has been well studied, little is known about the environmental diversity of antibiotic resistance genes (ARGs). ARGs and antibiotics have been the two aspects of life for microorganisms in the environment, even before they came to our attention and revolutionized medicine in the "antibiotic era". Today, culture-independent techniques such as metagenomics are beginning to reveal the diversity of antibiotic/antibiotic resistance systems found in the environment. Understanding antibiotic resistance mechanisms may have direct applications in the identification of new drug

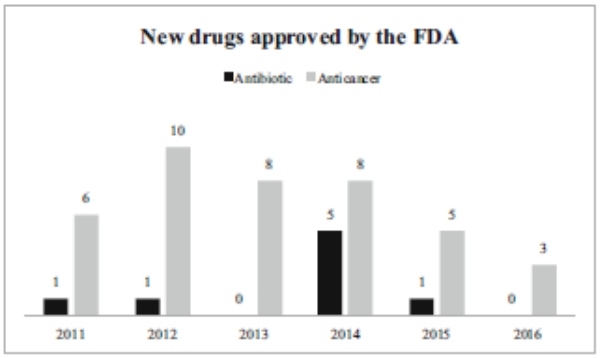


Fig. 1 New anticancer and antibiotic drugs approved by the Food and Drug Administration (FDA) since 2011. The *numbers above the bars* correspond to new formulations of drugs released on the market in the USA to treat specific infectious diseases and different types of cancer

targets and synthesis of novel and effective antibiotic molecules. Thus, studying ARGs is a welcome strategy in the war against microbial diseases.

The recently published work of Truman and collaborators provides an example of how resistance mechanisms can be used to identify new antimicrobials [2]. They exploited the glycopeptide resistance mechanism of *Streptomyces coelicolor* to develop a new two-step screen for antibiotic molecules (Fig. 2). In the first step, a cell wall stress assay is performed with a library of bacterial extracts to detect inducers of the *sigE* promoter, which is involved in the cell wall stress response. In the second step, extracts that activated the *sigE* promoter are tested using the Δ*femX* mutant bioassay, in which induction of glycopeptide resistance genes reveals the presence of a glycopeptide antibiotic. With this innovative strategy, they isolated an *Amycolatopsis* strain carrying a new 79-kb biosynthetic gene cluster for ristocetin, a previously uncharacterized glycopeptide antibiotic.

More recently, a new antimicrobial was identified using a culture-independent method that screens bacteria grown in diffusion chambers in situ. This technique enables the growth of bacteria by allowing them access to substrates from their natural environment. In this way, the authors isolated teixobactin, a powerful antibiotic that is active against clinically important Gram-positive bacteria. It is important to note that resistance mechanisms against this antibiotic were not found in *Staphylococcus aureus* or *Mycobacterium tuberculosis* strains. However, the authors acknowledge that resistant traits could be acquired by horizontal transfer [3].

Antibiotic biosynthetic clusters also encode ARGs, which provide "self-resistance" by various mechanisms, as reviewed by Allen et al. [3]. For instance, the biosynthetic cluster for the antibiotic actinorhodin contains genes encoding export proteins. Bacteria that do not produce antibiotics may acquire mutations in the genes encoding targets of antibiotics (e.g., DNA gyrase subunit A gene) that confer resistance. However, these mutations may have pleiotropic effects, altering metabolic pathways and consequently resulting in different levels of resistance.

Molecules that function as antibiotics at high concentrations appear to play different roles when present in low (and non-toxic) concentrations. The main role involves signaling and communication between bacteria, which may induce the expression of genes involved in virulence and other processes. In addition, antibiotics are an energy source for some bacteria, which may account for the high levels of antibiotic resistance determinants in natural environments [4].

Nevertheless, much remains to be understood about antibiotic resistance mechanisms, particularly in highly diverse environments such as soil. Soil is composed of particles of different size and composition, gradients of water, minerals, and organic compounds. Its heterogeneous structure creates a multitude of microenvironments that allow for extremely diverse microbial communities to exist. Soil is therefore an important reservoir of genetic and metabolic diversity, making it an ideal environment for the identification of novel ARGs, which can expand our knowledge of antibiotic resistance systems.

Over the past decade, the advent of metagenomics has vastly expanded our knowledge of microbiology. The shifts in focus from single isolates to microbial communities and from cultured to uncultured microorganisms have revolutionized the field. In fact, a new tree of life was recently proposed by Hug and collaborators [5], in which they proposed a new subdivision within the *Bacteria* domain called candidate phyla radiation (CPR). This new, higher-resolution tree was constructed by aligning ribosomal protein sequences rather than 16S rRNA genes. Members of the CPR group have relatively small genomes and limited metabolic capabilities, suggesting that many are symbionts. Interestingly, the lineages

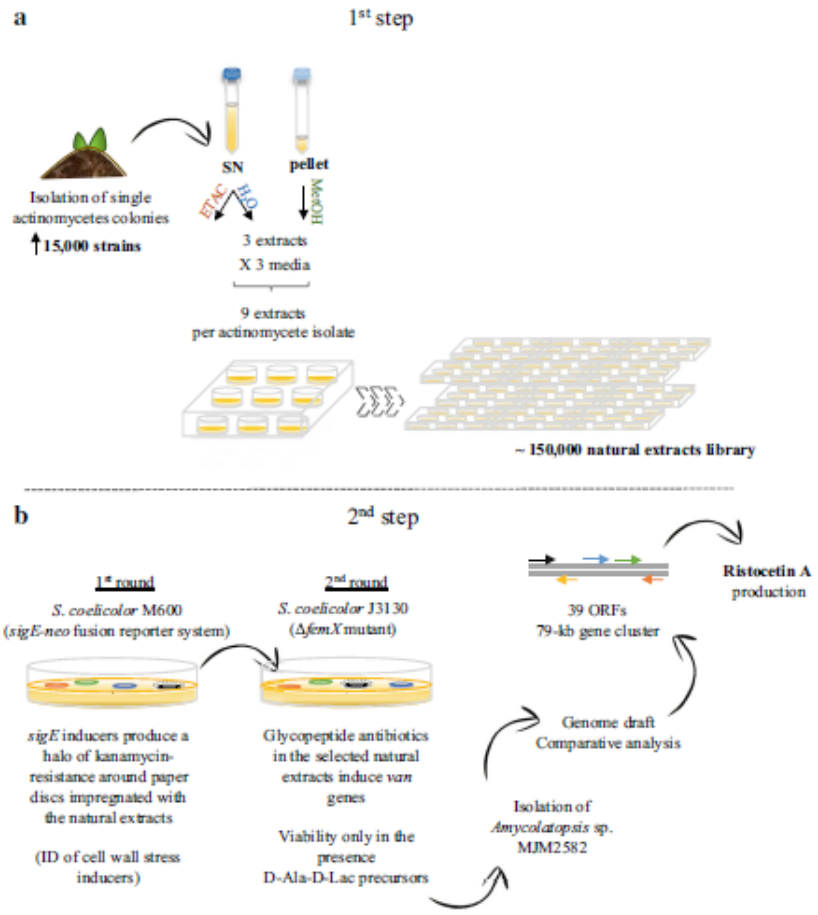


Fig. 2 Trumanet al. developed a two-step assay for the bioprospection of glycopeptide antibiotics by exploiting *van* gene-induced antibiotic resistance. **a** First, 15,000 actinomycete strains were isolated from a soil sample from south-eastern Asia. A natural product extract library was constructed with the aqueous and ethyl acetate layers of the culture supernatant and a methanol extraction of the cell pellet of each one of the previously isolated actinomycete strains, yielding an extract library of around 150,000 natural product extracts. **b** In the first round of screening, the extract library was screened using a *S. coelicolor* M600 strain carrying a *sigE-neo* fusion reporter system, which senses cell wall stress and expresses a kanamycin-dependent resistance phenotype. This reporter system identifies extracts that trigger a cell wall stress response. In the second round, extracts identified in the first round were screened using a *S. coelicolor* Δ*femX* mutant as follows. In the absence of vancomycin, peptidoglycan precursors with D-Ala-D-Ala chains are used for cell wall synthesis catalyzed by the FemX enzyme. However, when *van* genes are induced by vancomycin, D-Ala-D-Lac-containing precursors are produced and used by the homologous enzyme VanK to catalyze cell wall synthesis. This drug-dependent bioassay for glycopeptide bioprospection takes advantage of the inability of Δ*femX* mutants to grow unless *vanHAX* genes are expressed (i.e., glycopeptide antibiotics are present). The authors were able to isolate an *Amycolatopsis* strain carrying a 79-kb biosynthetic cluster for the glycopeptide antibiotic ristocetin

of CPR and other groups without cultivable members comprise most of the tree's diversity, highlighting the importance of culture-independent techniques for microbial diversity studies. Of course, these studies would not have been possible without efficient and inexpensive high-throughput next-generation sequencing (NGS).

Both sequence-based and functional metagenomic techniques can be used to assess the taxonomic and functional diversity of microorganisms from natural environments. Sequence-based metagenomics involves the direct sequencing of total DNA extracted from an environmental sample. This approach can be used to identify gene variants, describe functional and taxonomic diversity and abundance, and assemble microbial genomes. NGS technology has made sequence-based metagenomics popular among researchers and provided increased depth of sequencing, enabling the assembly of full genomes from metagenomic data. In addition, combining different sequencing technologies to obtain both short and long reads is more advantageous than using either technology alone [6]. For instance, Sharon et al. described the assembly of low-abundance genomes from complex environments in terrestrial sediments, which harbor many microorganisms belonging to candidate phyla or close to branches with no cultivable members. By using a combination of short- and long-read sequencing, they were able to demonstrate that microbial communities in this environment consist of numerous rare species but few abundant species.

On the other hand, functional metagenomics is a robust technique for identifying new biocatalysts of interest from diverse environmental samples. Functional metagenomics typically involves the construction of metagenomic libraries, followed by the selection of clones expressing the phenotype of interest, such as enzymatic or antimicrobial activity, antibiotic/xenobiotic resistance, or ability to degrade toxic compounds or environmental pollutants. The phenotypic-based screening of metagenomic clones provides invaluable information about gene function and is therefore used for gene annotation. Recent examples where this approach was successfully used include the identification of amylolytic, cellulolytic, and lipolytic activities with potential industrial applications [7–14]. Functional metagenomics is also useful for antimicrobial bioprospection [15] and bioremediation of soils contaminated with aromatic compounds [16, 17]. In ARG research, the functional metagenomics approach usually entails the selection of clones on agar plates supplemented with antibiotics, followed by analysis of their genetic content.

Soil microorganisms are a proven source of both antibiotics and ARGs. In fact, more than 80 % of all antibiotics used in the clinic are derived from soil microorganisms, which are, not surprisingly, also rich in antimicrobial resistance elements [18, 19]. Therefore, exploring the collection of ARGs from natural environments, the so-called resistome, may shed light on the true diversity of ARGs, identifying new resistance mechanisms and revealing their ecological roles. Objectives of ARG research include determining (1) whether antibiotic resistance is the primary function of these genes, (2) their connection to clinically important resistance mechanisms, and (3) their biotechnological potential.

Resistomes in Natural Environments

Several metagenomic-based studies have focused on characterizing ARGs from natural environments using functional or sequence-based approaches (Table 1). The studies analyzed soil, water, sludge, or a mixture of environmental samples from different sources to identify genes conferring resistance to a wide variety of antibiotics including amphenicols, aminoglycosides, β-lactams, fluoroquinolones, lincosamides, macrolides, multiple drugs (MDR), phosphonic acids, polypeptides, polymyxins, rifamycins, sulfonamides, tetracyclines, and trimethoprim.

One important difference between the sequence-based and functional metagenomic studies in Table 1 is that sequence-based approaches can identify only previously known ARGs through in silico analysis of metagenomic reads using ARG databases such as the Antibiotic Resistance Database (ARDB) [44] and Comprehensive Antibiotic Resistance Database (CARD) [45]. On the other hand, functional metagenomic surveys are able to identify new resistance mechanisms or assign new roles to previously known proteins by identifying and isolating potential open reading frames (ORFs) for phenotypic and protein characterization. For example, Amos and collaborators [39] screened a water sample for quinolone resistance and identified new roles for regulatory protein X (RecX) and recombinase A (RecA), which were not previously identified as ARGs. The authors provided evidence that the modulation of RecA by RecX may be involved in repairing DNA damage caused ciprofloxacin inhibition of DNA gyrase.

Two other examples of functional metagenomics studies that identified new ARGs include studies that screened soil metagenomic libraries for resistance to β-lactam antibiotics and identified a response regulator [31] and a dioxygenase [38]. In the first study, Allen and collaborators described a 5169-bp ORF encoding a response regulator gene derived from a resistant Alaskan soil metagenomic clone that also contained a putative metallopeptidase gene. Phenotypic tests showed that the response regulator gene was sufficient to confer resistance to carbenicillin, not by inactivating the antibiotic but rather by altering the expression of native *Escherichia coli* genes encoding porins and efflux pumps.

In the second study [38], the authors from this review screened a small-insert Brazilian Cerrado soil metagenomic library for resistance to the following nine β-lactam antibiotics: amoxicillin 16 μg mL^{-1}, ampicillin 50 μg mL^{-1}, carbenicillin 50 μg mL^{-1}, cephalexin 50 μg mL^{-1}, cefamandole 16 μg mL^{-1}, cefoxitin 20 μg mL^{-1}, ceftazidime 5 μg mL^{-1}, penicillin G 50 μg mL^{-1}, and piperacillin 12.5 μg mL^{-1}. After isolating resistant clones and subcloning genes of interest, a 1062-bp ORF that conferred carbenicillin resistance in *E. coli* was identified. Although gene annotation showed that the closest match was a hypothetical protein, hits included a bicupin domain and a putative gentisate 1,2-dioxygenase

Table 1 Metagenomic surveys for antibiotic resistance genes in natural environments

Approach	Sample	Antibiotic class																		Reference
Sequence		AG	AF	BL	FQ	GP	LS	MC	PA	PP	PY	RY	SF	TET	TMT	MDR	OTHER			
		Resistance determinants																		
	Soil	AAT	CAT	*ACT-1*					*floR*					*tet32*		*acrA*	AcRPB			[20]
		aadA	*catB2*	*ampC*					FRP					*tetA*		*acrB*	AFRPB			
		A3PH	*catB3*	*CARB*					*rosA*					*tetA(31)*		*adeA*	AEP			
		aphD	*cmlA*	*L1*										*tetA(33)*		*adeB*	AcRPA			
		ARPA	CFE	*mecA*										*tetA(39)*		*AmrB*	AFRPA			
		KNT	CFRP	*OXA*										*tetA(41)*		*bpeF*				
		FAT	*cmlA*	*PSE-1*										*tetA(B)*		*emrFS*				
		HARK	*cmx*	*SedI*										*tetA(D)*		*emrG*				
		APT	CRD	*TEM-21*										*tetA(G)*		*emrGM*				
		ARPB	CT	*VIM-18*										*tetA(M)*		*emrT*				
		aac6	*floR*											*tetA(P)*		*emrX*				
		aadB												*tetB*		*mdfA*				
		GAT												*tetG*		*mdfE*				
		GRP												*tetH*		*mdtH*				
		aac												*tetL*		*mdtL*				
		ere												*tetM*		*mdtO*				
														tetO		*mef*				
														tetQ		*mexA*				
														tetS		*mexB*				
														tetT		*mexD*				
														tetV		*mexE*				
														tetW		*mexF*				
														tetX		*mexG*				
														tetX2		MLSBP				
														tetY		*mtrA*				
																qacH				
																ftpa				
																tdeT				
	Water												*sul1*	*tet33*		*acrB*				[21]
													sul2	*tet34*		*mexB*				
														tetC		*mexF*				
														tetL						
														tetM						
														tetX						
														tetL						[22]
	Sludge												*sul1*	*tet32*	*dfrA7*	*acrB*				[23]
													sul2	*tet33*		*emrB*				
														tetA		*emrC*				
														tetC		*emrQ*				
														tetG		*emrT*				
														tetO		*mdtF*				
														tetPa		*mexB*				
														tetPb		*mexD*				
														tetV		*mexF*				
														tetW		*mexW*				

Table 1 (continued)

Approach	Sample	Antibiotic class															Reference
Sequence		AG	AF	BL	FQ	GP	LS	MC	PA	PP	PY	RY	SF	TET	TMT	MDR	OTHER	
		Resistance determinants																
														tetX		*macY*		
		aac		Class A BL										*tet39*				[24]
		aadA		*NPS-1*										*tetA*				
		AGRPA		*OXA-1*										*tetC*				
		aph		*OXA-10*										*tetG*				
				OXA-2										*tetM*				
				VEB-5										*tetO*				
														tetP				
														tetS				
														tetW				
														tetX				
		aadA	*catB*	*blaCFX-a*	*rmeB*			*macB*		*bacA*			*nul1*	*tet33*		*acrB*		[25]
		aadB	*cmlA*	*blaOXA-10*	*rmeE*								*nul2*	*tet36*		*adeA*		
		aphA-2	*floR*	*blaVEB-5*										*tet39*		*ermB*		
		strA		*blaVEB-3*										*tet40*		*mdtF*		
		strB												*tetA(P)*		*mexB*		
														tetB(P)		*mexF*		
														tetG		*mexW*		
														tetM				
														tetO				
														tetW				
	Mix	*aadA*	*catA2*	BL	*qnrB*	*vanA*	LNT	*ereA*	*rosA*		*arnA*		*nul1*	*tet32*	*dfrA*	*acrA*	*VatE*	[26]
		AAT	*catB2*	*CARB-8*	*qnrS*			*ereB*					*nul2*	*tet33*		*acrB*	*VgaB*	
		AAdt	*catB3*	*QnrA2*	QRP			ERM						*tet39*		*bcr*		
		APT	*catD*	*QnrA3*				ERP						*tetA*		*ermA*		
		ANT	CFE	*CMY-36*				*macB*						*tetC*		*ermB*		
		emrE	CRP	*CTX-M-1*										*tetG*		ermD		
			cmlA1	*CTX-M-14*										*tetH*		*ermF*		
			cmlA5	*CTX-M-22*										*tetL*		*ermG*		
			cmlA9	*CTX-M-32*										*tetM*		*ermT*		
			cmxA	*NPS-1*										*tetO*		*ermX*		
			floR	*OXA-1*										*tetP(A)*		HAE		
				OXA-10										*tetP(B)*		*mdtA*		
				OXA-9										*tetQ*		*mdtE*		
				OXA-2										*tetS*		*mdtF*		
				PBP										*tetT*		*mdtG*		
				PER-1										*tetV*		*mdtH*		
				ROB-1										*tetW*		*mdtK*		
				TEM										*tetX*		*mdtL*		
				VEB-1										*tetY*		*mdtM*		
				VEB-5										TRP		*mdtN*		
														TEP		*mdtO*		
																mdtP		
																mexB		
																mexF		
																MFS		
																MLS		

Table 1 (continued)

Approach	Sample	Antibiotic class																Reference
Sequence		AG Resistance determinants	AF	BL	FQ	GP	LS	MC	PA	PP	PY	RY	SF	TET	TMT	MDR	OTHER	
																norM PT2 RND SAT *sdeY* *tolc*		
		2-NAT A3PH A6PH *aadA* AAT AGRPA AGRPB	CAT CFE *cmlA9* FE	BL Class A BL *NPS-1* *OXA* *OXA-10* *OXA-46* PBP *VEB-1* *VEB-3* *VIM-11*	QRP			EE ERM *macB* McT *mef* *mefA* MEP		*strB* UD UK	*arnA*		*sul1* *sul2*	*tet32* *tetA* *tetA(33)* *tetA(C)* *tetA(M)* *tetA(Q)* *tetB(P)* *tetG* *tetM* *tetO* *tetQ* *tetV* *tetW* *tetX* TRP		ABC *acrB* *emrB* *emrF* HAE *mdtF* *mdtG* *mdtH* *mdtN* *mexB* *mexD* *mexF* *norM* RND *sdeY*	CPA2 PAP UDP	[27]
		ksgA	*catA* *ceo* *cml*	*blaA* *blaB* *pbp*	*smeA* *smeB* *smeC* *smeD* *smeE* *smeF*	*vanA* *vanB* *vanC* *vanD* *vanE* *vanG*		*macA* *macB*	*rosA* *rosB*	*bacA* *bcr*	*arnA*		*sul*	*tet (efflux)* *tet (flavo)* *tet (rpp)*	*dfrA*	*acr* *adeA* *adeB* *adeC* *amr* *mdtE* *mdtF* *mdtK* *mdtN* *mdtO* *mdtP* *mepA* *mexA* *mexB* *mexC* *mexD* *mexE* *mexF* *mexH* *mexI* *mexV* *mexW* *mexX* *mexY* *norM* *qac*	*pur8* *tcmA* *vat*	[28]

Table 1 (continued)

Approuch	Sample	Antibiotic class																		Reference
Sequence		AG	AF	BL	FQ	GP	LS	MC	PA	PP	PY	RY	SF	TET	TMT	MDR	OTHER			
		Resistance determinants																		
		30S-S5	*aar*	OXL	*gyrA*	*vanR*		50S-L17				rpoB	*sul1*	*tet56*		ABC	AEP		[29]	
		30S-S12	*cml*	PBP	*parC*	*vanS*		*macB*					*sul2*	*tetA*		*ozrB*	DNA-TI-IV			
		AAdt	*flo*					MdT						*tetC*		*ermLF*	*emhB*			
		ANT												*tetG*		*erm(G)*	*gyrB*			
		APT												*tetM*		*erm(TR)*				
														tetO		*mdtB*				
														tetW		*mdtC*				
														tetX		MFS				
																RND				
		aac6ib			*qnpA*														[30]	
					qnrA															
					qnrB															
					qnrS															
	Functional	Soil		Response regulator															[31]	
									MurA										[32]	
		3-NAT										R-ADPRT							[33]	
		6-NAT										RMO								
		AAT		Class A BL										EP					[34]	
				Class C BL																
				Class D BL																
				Metallo BL																
				Bifunctional															[35]	
				BL																
		MT		BL												MFS			[36]	
																ABC				
																PT				
																AT				
																TS				
		2-NAT	ANA											AL		ABC			[37]	
		2-OXO-DH	AL											STK		RND				
		3-NAT	ATPp											*orfC*						
		6-NAT	CFE											*tet(34)*						
		AAdT	RP											*tet(37)*						
		ATD1	STK											*tetV*						
		EEL												*tetX*						
		FAA																		
		GBD																		
		HAD																		
		ICMT																		
		ML																		
		PM																		
		NSDH																		
		TL																		
		tRNAs																		
				1,2-GDO															[38]	
	Water	K+ transport prn		TEM BL	RecX														[39]	
		3-NAT			RA															

Table 1 (continued)

Approach	Sample	Antibiotic class																Reference
Sequence		AG	AF	BL	FQ	GP	LS	MC	PA	PP	PY	RY	SF	TET	TMT	MDR	OTHER	
		Resistance determinants																
		PDH GNAT AAdt A3PH 6-NAT MetRS GT																
			BcrCflA															[40]
				Class A BL														[41]
				Class A BL														[42]
	Sludge	OMT		Class C BL														[43]
		A3PH		TEM BL														
	Mix																	n/a

AF amphenicol, *AG* aminoglycoside, *BL* β-lactam, *FQ* fluoroquinolone, *LS* lincosamide, *MC* macrolide, *MDR* multidrug, *PA* phosphonic acid, *PP* polypeptide, *PY* polymyxin, *RY* rifamycin, *SF* sulfonamide, *TET* tetracycline, *TMP* trimethoprim, *1,2-GDO* gentisate 1,2-dioxygenase, *2-OXO-DH* 2-oxoglutarate dehydrogenase, *2-NAT* 2-N′-acetyltransferase, *3-NAT* 3-N′-acetyltransferase, *6-NAT* 6-N′-acetyltransferase, *30S-S5* spectinomycin-30S ribosomal protein S5, *30S-S12* streptomycin-30S ribosomal protein S12, *50S-L17* macrolide-50S ribosomal protein L17, *A3PH* aminoglycoside 3′-phosphotransferase, *A6PH* aminoglycoside 6′-phosphotransferase, *AAdT* aminoglycoside adenyltransferase, *AAT* aminoglycoside acetyltransferase, *ABC* ATP-binding cassette, *AcRPA* acridine resistance protein A, *AcRPB* acridine resistance protein B, *AEP* acridine efflux pump, *AFRPA* acriflavine resistance protein A, *AFRPB* acriflavine resistance protein B, *AGRPA* aminoglycoside resistance protein A, *AH* aminohydrolase-like protein, *AL* allantoinase-like protein, *ANA* alpha-N-arabinofuranosidase-like protein, *ANT* aminoglycoside nucleotidyltransferase, *APT* aminoglycoside phosphotransferase, *ARPA* aminoglycoside resistance protein A, *ARPB* aminoglycoside resistance protein B, *AT* acetyltransferase, *ATIB* aminotransfemse III, *ATPp* ATP phosphoribosyltransferase-like protein, *BcrCflA* multidrug resistance protein from BcrCflA subfamily, *BL* β-lactamase, *CAT* chloramphenicol acetyltransferase, *CFE* chloramphenicol and florfenicol exporter, *CFRP* chloramphenicol and florfenicol resistance protein, *CPA2* CPA2 family, *CRD* chloramphenicol resistance determinant, *CT* chloramphenicol transporter, *DNA-TI-IV* DNA topoisomerase IV subunit beta, *EE* erythromycin esterase, *EEL* erythromycin esterase-like protein, *EP* efflux pump, *ERM* erythromycin ribosome methylase, *ERP* erythromycin resistance protein, *FAA* 2-fumarylacetoacetate, *FAT* fused aminoglycoside 3′-adenyltransferase-aminoglycoside resistance protein B, *FE* florfenicol exporter, *FRPB* fosfomycin resistance protein, *GAT* gentamycin acetyltransferase, *GBD* glyoxalase/bleomycin resistance/dioxygenase, *GRP* gentamycin resistance protein, *GNAT* acetyltransferase family protein, *GT* glycosyltransferase 25 family member 1, *HAD* HAD family hydrolase-like protein, *HAE* HAE1 family protein, *HARK* hydroxyurea antibiotic resistance kinase, *LNT* lincosamide nucleotidyltransferase, *ICMT* isoprenylcysteine carboxyl methyltransferase, *KNT* kanamycin nucleotidyltransferase, *MCT* macrolide transporter, *MEP* macrolide efflux protein, *MLSRP* macrolide-lincosamide-streptogramin resistance protein, *MetRS* methionyl-tRNA synthetase, *MFS* major facilitator superfamily, *ML* methylase-like protein, *MLS* MLS-B resistance protein, *MurA* fosfomycin resistant protein, *MT* methyltransferase, *NSDH* nucleotide sugar dehydrogenase, *OMT* O-methyltransferase, *OXL* oxacillinase, *PAP* proton antiport protein, *PBP* penicillin binding protein, *PDH* pyruvate dehydrogenase subunit E1, *PM* phosphomannomutase, *PT* phosphotransfemse, *PT2* phosphotransferase type 2, *QRP* quinolone resistance protein, *R-ADP-RT* rifampin ADP-ribosyltransfemse, *RA* recombinase A, *RecX* regulatory protein X, *RND* RND multidrug efflux pump, *RMO* rifampin monooxygenase, *RND* resistance-nodulation-cell division efflux pump, *RP* ribosome protection protein, *rpoB* DNA-directed RNA polymerase subunit beta, *SAT* streptogramin-A-acetyltransfemse, *STK* serine/threonine protein kinase-like, *TL* transketolase-like protein, *tRNAs* tRNA synthetase, *TEP* tetracycline efflux protein, *TRP* tetracycline resistance protein, *TS* thymidylate synthase, *UD* undecaprenyl diphosphatase, *UDP* UDP-glucuronic acid decarboxylase/UDP-4-amino-4-eoxy-L-arabinose formyltransferase, *UK* undecaprenol kinase

(GDO). Additional functional and bioinformatics analysis characterized the metagenomic insert as a new GDO. This enzyme also conferred resistance to phenol, but the mechanism of action for both resistance phenotypes requires further investigation. The enzyme may also have biotechnological potential beyond human health, given this activity against aromatic compounds.

In the literature and in sequence databases, genes encoding β-lactamases are typically associated with clones resistant to β-lactam antibiotics. Nevertheless, the two studies cited above [31, 38] demonstrate that ARG research can reveal new and unexpected antimicrobial resistance mechanisms and provide insight into the functions of ARGs in natural environments. It should be emphasized that, in both studies, not only did the identified genes not encode β-lactamases but also neither of these ORFs showed similarity to resistance elements in public sequence databases. These results indicate that the diversity of ARGs in natural environments is much greater than previously thought, and efforts must be made to establish additional strategies for identifying new ARGs in the environment.

Table 1 also shows examples of novel members of previously known ARG families. They include UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) with a mutated active site that confers fosfomycin resistance [32]; a bifunctional β-lactamase, reported as a natural fusion of class C and D enzymes, which is twice the size of previously characterized members [35]; two class A β-lactamases [41, 42]; a dihydrofolate reductase and a rifampin ADP-ribosyltransferase [40]; and a novel trimethoprim-resistant reductase [19].

Although the functional metagenomics approach has the advantage of not being limited to previously known sequences, some ARGs are expressed at low levels or not expressed at all (i.e., cryptic resistance genes), which is an obstacle for their identification by phenotypic screens. Inefficient gene expression caused by codon bias or problems with promoter recognition in the host organism can be addressed in two ways. Metagenomic libraries can be constructed using broad-host-range vectors and a range of bacterial hosts. Because codon usage bias varies across species, each host offers the opportunity for the expression of a different set of genes. For example, Craig et al. reported little overlap of positive clones in a functional metagenomics study using different *Proteobacteria* host species. The clones were screened for antibacterial activity, pigmentation changes, and colony morphology. Only one small molecule was common to both metagenomic libraries established using the two different species, showing the power of this strategy to retrieve different active clones and expand the repertoire of bioactive molecules identified in functional metagenomic screens [46].

The second technique to improve gene expression in clones involves the exchange of native promoters that are silent in a particular host for new promoters. A recently published proof-of-concept paper [47] described the use of the CRISPR/Cas9 system to induce site-specific double-strand breaks in promoter DNA. This technique increased recombination frequency in yeast, enabling the marker-free replacement of promoters by transformation-associated recombination. It remains to be seen how useful this technique will be for antibiotic discovery because only one gene cluster can be studied at a time, and the expressed molecule will still needs to be tested for antibiotic activity.

It is worth noting that at some point even functional-based metagenomic studies require the use of sequence databases. However, the functional roles of all peptides encoded by the ORF are considered when searching for new ARGs, rather than focusing only on previously known ARG sequences. After a clone is identified as having a specific activity and sequenced, identification of the potential ARGs can pose a challenge, particularly if similarity to genes in databases does not implicate a particular ORF as the gene responsible for the resistance phenotype. For instance, ARG can be characterized through mutagenesis [31, 32, 34, 35, 39, 41, 43], construction of sub-libraries [30, 36, 42], or identified by similarities in sequence databases [25, 37, 45, 48].

Because the successful characterization of novel ARGs first requires the identification of ORFs of interest, it is important to consider insert size in the metagenomic library. Libraries with inserts no larger than 10 kb are less time-consuming and easier to work with. However, small-insert libraries often allow only for the identification of single genes able to confer resistance by themselves in a particular host. On the other hand, large-insert libraries can provide information about the context of a given ARG, such as whether it is part of an operon or gene cluster, or located on a mobile element. In addition, large inserts can provide information about adjacent genes. For that reason, combining data from small- and large-insert libraries may be a useful strategy to identify and characterize novel ARGs.

Identifying the genomic location of an ARG can reveal new gene cassettes, such as integrons, the most recently discovered mechanisms of horizontal gene transfer between microorganisms [49]. Integrons consist of a promoter, recombination site, and gene encoding a site-specific recombinase (integrase), which is able to capture and incorporate external DNA into the cassette. Hence, integrated genes are found in tandem. Integrons lack self-mobility, but they can be found on plasmids and transposons, which enable their movement [48, 50]. Because chromosomal integrons are hot spots for recombination and chromosomal alterations, such as rearrangements, they generate substantial diversity even within a species. Chromosomal integrons appear to be more common in environmental bacteria, whereas clinical strains are more likely to harbor plasmid- or transposon-borne integrons [48]. The ability of integrons to capture diverse genes makes them important elements for microbial adaptation and evolution, particularly when multidrug resistance is concerned.

There are five classes of integrons associated with the antibiotic resistance phenotype, many already associated with clinical isolates [48]. In this respect, class 1 integrons are one of the major structures on the spread of antimicrobial resistance genes. The selective pressure of extensive antibiotic and other toxic substances usage, which makes their evolutionary history very recent, impelled the evolution of antibiotic-resistance integrons, mainly from classes 1 to 3. In this regard, it is thought that class 1 integrons are derived from chromosomal integrons found in environmental isolates, which today continues to assimilate new cassettes containing ARGs and other genes encoding different phenotypes. This dissemination shows the efficacy of natural selection in the evolution of multiresistant microorganisms, specially related to clinical and pathogenic strains [48, 50].

Exploring their environmental diversity is a useful strategy to predict, and ultimately prevent, their transference to clinically important microorganisms.

Conclusions

Antimicrobial resistance is a major threat to human health; therefore, a better understanding of resistance mechanisms is needed to control clinically important multidrug-resistant strains. Although antibiotic resistance in nosocomial strains is relatively well understood, the natural environment contains a vast and largely unexplored diversity of ARGs. Culture-independent metagenomic approaches have proven to be invaluable tools to access this genetic wealth. The identification and description of new resistance mechanisms may facilitate the bioprospection of novel antimicrobial compounds and development of new drugs. The addition of novel resistance elements to ARG databases, such as ARDB and CARD, is essential to support further studies on antimicrobial resistance and the roles of ARGs outside the clinical setting.

In this review, we described recent studies that highlight the broad diversity of ARGs in natural environments, especially in soils. Functional metagenomics is an important approach in ARG research, since it can be used to identify and characterize new ARGs, including those not previously associated with antibiotic resistance. This strategy will give us the much-needed leverage to combat multiple drug-resistant microorganisms of clinical relevance and to describe the ecological aspects of new environmental ARGs.

Acknowledgments The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support.

References

1. Seiple IB, Zhang Z, Jakubec P, Langlois-Mercier A, Wright PM, Hog DT, Yabu K, Allu SR, Fukuzaki T, Carlsen PN, Kitamura Y, Zhou X, Condakes ML, Szczypinski FT, Green WD, Myers AG (2016) A platform for the discovery of new macrolide antibiotics. Nature 533:338–345. doi:10.1038/nature17967
2. Truman AW, Kwun MJ, Cheng J, Yang SH, Suh JW, Hong HJ (2014) Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel amycolatopsis strain producing ristocetin. Antimicrob Agents Chemother 58:5687–5695. doi:10.1128/AAC.03349-14
3. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schaberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459. doi:10.1038/nature14098
4. Martinez JL (2008) Antibiotics and antibiotic resistance genes in natural environments. Science 321:365–367. doi:10.1126/science.1159483
5. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the tree of life. Nature Microbiology 1. doi: 10.1038/nmicrobiol.2016.48
6. Sharon I, Kertesz M, Hug LA, Pushkarev D, Blauwkamp TA, Castelle CJ, Amirebrahimi M, Thomas BC, Burstein D, Tringe SG, Williams KH, Banfield JF (2015) Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. Genome Res 25:534–543. doi:10.1101/gr.183012.114
7. Pereira MR, Mercaldi GF, Maester TC, Balan A, de Macedo Lemos EG (2015) Est16, a new esterase isolated from a metagenomic library of a microbial consortium specializing in diesel oil degradation. PLoS One 10:e0133723. doi:10.1371/journal.pone.0133723
8. Prive F, Newbold CJ, Kadeebhai NN, Girdwood SG, Golyshina OV, Golyshin PN, Scollan ND, Huws SA (2015) Isolation and characterization of novel lipases/esterases from a bovine rumen metagenome. Appl Microbiol Biotechnol 99:5475–5485. doi:10.1007/s00253-014-6355-6
9. Kim HJ, Jeong YS, Jung WK, Kim SK, Lee HW, Kahng HY, Kim J, Kim H (2015) Characterization of novel family IV esterase and family L3 lipase from an oil-polluted mud flat metagenome. Mol Biotechnol 57:781–792. doi:10.1007/s12033-015-9871-4
10. Su J, Zhang F, Sun W, Karuppiah V, Zhang G, Li Z, Jiang Q (2015) A new alkaline lipase obtained from the metagenome of marine sponge *Ircinia* sp. World J Microbiol Biotechnol 31:1093–1102. doi:10.1007/s11274-015-1859-5
11. Alnoch RC, Martini VP, Glogauer A, Costa AC, Piovan L, Muller-Santos M, de Souza EM, de Oliveira Pedrosa F, Mitchell DA, Krieger N (2015) Immobilization and characterization of a new regioselective and enantioselective lipase obtained from a metagenomic library. PLoS One 10:e0114945. doi:10.1371/journal.pone.0114945
12. Vester JK, Glaring MA, Stougaard P (2015) An exceptionally cold-adapted alpha-amylase from a metagenomic library of a cold and alkaline environment. Appl Microbiol Biotechnol 99:717–727. doi:10.1007/s00253-014-5931-0
13. Xu B, Yang F, Xiong C, Li J, Tang X, Zhou J, Xie Z, Ding J, Yang Y, Huang Z (2014) Cloning and characterization of a novel alpha-amylase from a fecal microbial metagenome. J Microbiol Biotechnol 24:447–452
14. Kanokratana P, Eurwilaichitr L, Pootanakit K, Champreda V (2015) Identification of glycosyl hydrolases from a metagenomic library of microflora in sugarcane bagasse collection site and their cooperative

action on cellulose degradation. J Biosci Bioeng 119:384–391. doi:10.1016/j.jbiosc.2014.09.010
15. O'Mahony MM, Henneberger R, Selvin J, Kennedy J, Doohan F, Marchesi JR, Dobson AD (2015) Inhibition of the growth of *Bacillus subtilis* DSM10 by a newly discovered antibacterial protein from the soil metagenome. Bioengineered 6:89–98. doi:10.1080/21655979.2015.1018493
16. Nagayama H, Sugawara T, Endo R, Ono A, Kato H, Ohtsubo Y, Nagata Y, Tsuda M (2015) Isolation of oxygenase genes for indigo-forming activity from an artificially polluted soil metagenome by functional screening using *Pseudomonas putida* strains as hosts. Appl Microbiol Biotechnol 99:4453–4470. doi:10.1007/s00253-014-6322-2
17. Lee CM, Yeo YS, Lee JH, Kim SJ, Kim JB, Han NS, Koo BS, Yoon SH (2008) Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome. Biochem Biophys Res Commun 370:322–326. doi:10.1016/j.bbrc.2008.03.102
18. D'Costa VM, Griffiths E, Wright GD (2007) Expanding the soil antibiotic resistome: exploring environmental diversity. Curr Opin Microbiol 10:481–489. doi:10.1016/j.mib.2007.08.009
19. Torres-Cortes G, Millan V, Ramirez-Saad HC, Nisa-Martinez R, Toro N, Martinez-Abarca F (2011) Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. Environ Microbiol 13:1101–1114. doi:10.1111/j.1462-2920.2010.02422.x
20. Fang H, Wang H, Cai L, Yu Y (2015) Prevalence of antibiotic resistance genes and bacterial pathogens in long-term manured greenhouse soils as revealed by metagenomic survey. Environ Sci Technol 49:1095–1104. doi:10.1021/es504157v
21. Wang Z, Zhang XX, Huang K, Miao Y, Shi P, Liu B, Long C, Li A (2013) Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. PLoS One 8:e76079. doi:10.1371/journal.pone.0076079
22. Yang J, Wang C, Shu C, Liu L, Geng J, Hu S, Feng J (2013) Marine sediment bacteria harbor antibiotic resistance genes highly similar to those found in human pathogens. Microb Ecol 65:975–981. doi:10.1007/s00248-013-0187-2
23. Huang K, Tang J, Zhang XX, Xu K, Ren H (2014) A comprehensive insight into tetracycline resistant bacteria and antibiotic resistance genes in activated sludge using next-generation sequencing. Int J Mol Sci 15:10083–10100. doi:10.3390/ijms150610083
24. Yang Y, Li B, Ju F, Zhang T (2013) Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. Environ Sci Technol 47:10197–10205. doi:10.1021/es4017365
25. Zhang T, Zhang XX, Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. PLoS One 6:e26041. doi:10.1371/journal.pone.0026041
26. Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, Zhang T (2015) Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J 9:2490–2502. doi:10.1038/ismej.2015.59
27. Yang Y, Li B, Zou S, Fang HH, Zhang T (2014) Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. Water Res 62:97–106. doi:10.1016/j.watres.2014.05.019
28. Nesme J, Cecillon S, Delmont TO, Monier JM, Vogel TM, Simonet P (2014) Large-scale metagenomic-based study of antibiotic resistance in the environment. Curr Biol 24:1096–1100. doi:10.1016/j.cub.2014.03.036
29. Ma L, Li B, Zhang T (2014) Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis. Appl Microbiol Biotechnol 98:5195–5204. doi:10.1007/s00253-014-5511-3
30. Cummings DE, Archer KF, Arriola DJ, Baker PA, Faucett KG, Laroya JB, Pfeil KL, Ryan CR, Ryan KR, Zuill DE (2011) Broad dissemination of plasmid-mediated quinolone resistance genes in sediments of two urban coastal wetlands. Environ Sci Technol 45:447–454. doi:10.1021/es1029206
31. Allen HK, An R, Handelsman J, Moe LA (2015) A response regulator from a soil metagenome enhances resistance to the beta-lactam antibiotic carbenicillin in *Escherichia coli*. PLoS One 10:e0120094. doi:10.1371/journal.pone.0120094
32. Cheng G, Hu Y, Lu N, Li J, Wang Z, Chen Q, Zhu B (2013) Identification of a novel fosfomycin-resistant UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) from a soil metagenome. Biotechnol Lett 35:273–278. doi:10.1007/s10529-012-1074-5
33. McGarvey KM, Queitsch K, Fields S (2012) Wide variation in antibiotic resistance proteins identified by functional metagenomic screening of a soil DNA library. Appl Environ Microbiol 78:1708–1714. doi:10.1128/AEM.06759-11
34. Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein FC, McManus PS, Handelsman J (2010) Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. Appl Environ Microbiol 76:4396–4401. doi:10.1128/AEM.01763-09
35. Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. ISME J 3:243–251. doi:10.1038/ismej.2008.86
36. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G (2014) Bacterial phylogeny structures soil resistomes across habitats. Nature 509:612–616. doi:10.1038/nature13377
37. Su JQ, Wei B, Xu CY, Qiao M, Zhu YG (2014) Functional metagenomic characterization of antibiotic resistance genes in agricultural soils from China. Environ Int 65:9–15. doi:10.1016/j.envint.2013.12.010
38. dos Santos DF, Istvan P, Noronha EF, Quirino BF, Kruger RH (2015) New dioxygenase from metagenomic library from Brazilian soil: insights into antibiotic resistance and bioremediation. Biotechnol Lett 37:1809–1817. doi:10.1007/s10529-015-1861-x
39. Amos GC, Zhang L, Hawkey PM, Gaze WH, Wellington EM (2014) Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. Vet Microbiol 171:441–447. doi:10.1016/j.vetmic.2014.02.017
40. Lopez-Perez M, Mirete S, Jardon-Valadez E, Gonzalez-Pastor JE (2013) Identification and modeling of a novel chloramphenicol resistance protein detected by functional metagenomics in a wetland of Lerma, Mexico. Int Microbiol 16:103–111
41. Vercammen K, Garcia-Armisen T, Goeders N, Van Melderen L, Bodilis J, Cornelis P (2013) Identification of a metagenomic gene cluster containing a new class A beta-lactamase and toxin-antitoxin systems. Microbiologyopen 2:674–683. doi:10.1002/mbo3.104
42. Uyaguari MI, Fichot EB, Scott GI, Norman RS (2011) Characterization and quantitation of a novel beta-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. Appl Environ Microbiol 77:8226–8233. doi:10.1128/AEM.02732-10
43. Parsley LC, Consuegra EJ, Kakirde KS, Land AM, Harper WF Jr, Liles MR (2010) Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. Appl Environ Microbiol 76:3753–3757. doi:10.1128/AEM.03080-09
44. Liu B, Pop M (2009) ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443–447. doi:10.1093/nar/gkn656
45. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski

AC, Piddock LJ, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD (2013) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. doi:10.1128/AAC.00419-13

46. Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol 76: 1633–1641. doi:10.1128/AEM.02169-09

47. Kang HS, Charlop-Powers Z, Brady SF (2016) Multiplexed CRISPR/Cas9- and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. ACS Synth Biol. doi:10.1021/acssynbio.6b00080

48. Gillings MR (2014) Integrons: past, present, and future. Microbiol Mol Biol Rev 78:257–277. doi:10.1128/MMBR.00056-13

49. Stokes HW, Hall RM (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol Microbiol 3:1669–1683

50. Deng Y, Bao X, Ji L, Chen L, Liu J, Miao J, Chen D, Bian H, Li Y, Yu G (2015) Resistance integrons: class 1, 2 and 3 integrons. Ann Clin Microbiol Antimicrob 14:45. doi:10.1186/s12941-015-0100-6

REFERENCES

Adler, A. J., Greenfield, N. J., and Fasman, G. D. (1973). Circular dichroismus and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol.* 27, 675–735.

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Arpigny, J. L., and Jaeger, K. E. (1999). Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* 343 Pt 1, 177–183.

Arpigny, J. L., Jendrossek, D., and Jaeger, K. E. (1998). A novel heat-stable lipolytic enzyme from Sulfolobus acidocaldarius DSM 639 displaying similarity to polyhydroxyalkanoate depolymerases. *FEMS Microbiol. Lett.* 167, 69–73. doi:10.1016/S0378-1097(98)00375-9.

Bairoch, A., and Boeckmann, B. (1994). The SWISS-PROT protein sequence data bank: current status. *Nucleic Acids Res.* 22, 3578–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308324&tool=pmcentrez&rendertype=abstract.

Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017). Metagenomics: Novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* 364. doi:10.1093/femsle/fnx211.

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235.

Bhattacharya, D., and Cheng, J. (2013). 3Drefine: consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. *Proteins* 81, 119–31. doi:10.1002/prot.24167.

Böhm, G., Muhr, R., and Jaeinicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* 5, 191–195.

Bonelli, F. S., and Jonas, a (1993). Reaction of lecithin: cholesterol acyltransferase with a water soluble substrate: effects of surfactants. *Biochim. Biophys. Acta* 1166, 92–98. doi:0005-2760(93)90288-K [pii].

Bornscheuer, U. T. (2002). Microbial carboxyl esterases: Classification, properties and application in biocatalysis. *FEMS Microbiol. Rev.* 26, 73–81. doi:10.1016/S0168-6445(01)00075-4.

Borrelli, G. M., and Trono, D. (2015). Recombinant lipases and phospholipases and their use as biocatalysts for industrial applications. *Int. J. Mol. Sci.* 16, 20774–20840. doi:10.3390/ijms160920774.

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Choo, D., Kurihara, T., Suzuki, T., and Soda, K. (1998). A Cold-Adapted Lipase of an Alaskan Gene Cloning and Enzyme Purification and Characterization. *Appl. Environ. Microbiol.* 64, 1–7.

Choudhury, P., and Bhunia, B. (2015a). Industrial Application of Lipase: a Review. *Biopharm J.* 1, 41–47. Available at: http://www.biopharmj.com/journal/index.php/BIOPHARMJ/article/view/11.

Choudhury, P., and Bhunia, B. (2015b). Industrial Application of Lipase: a Review. *Biopharm J.* 1, 41–47. Available at: http://www.biopharmj.com/journal/index.php/BIOPHARMJ/article/view/11.

Cygler, M., and Schrag, J. D. (1999). Structure and conformational flexibility of Candida rugosa lipase. *Biochim. Biophys. Acta* 1441, 205–14. doi:10.1016/S1388-1981(99)00152-3.

de Castro, A. P., Quirino, B. F., Allen, H., Williamson, L. L., Handelsman, J., and Krüger, R. H. (2011). Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content. *Biotechnol. Lett.* 33, 2169–2175. doi:10.1007/s10529-011-0693-6.

De Godoy Daiha, K., Angeli, R., De Oliveira, S. D., and Almeida, R. V. (2015). Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *PLoS One* 10, 1–20. doi:10.1371/journal.pone.0131624.

DeLano, W. (2008). The PyMOL Molecular Graphics System.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol.* 77, 1315–1324. doi:10.1128/AEM.01526-10.

Dewan, S. S. (2017). Global Markets for Enzymes in Industrial Applications. Available at: www.bccresearch.com.

dos Santos, D. F. K., Istvan, P., Noronha, E. F., Quirino, B. F., and Krüger, R. H. (2015). New dioxygenase from metagenomic library from Brazilian soil: insights into antibiotic resistance and bioremediation. *Biotechnol. Lett.* 37, 1809–1817. doi:10.1007/s10529-015-1861-x.

dos Santos, D. F. K., Istvan, P., Quirino, B. F., and Kruger, R. H. (2017). Functional Metagenomics as a Tool for Identification of New Antibiotic Resistance Genes from Natural Environments. *Microb. Ecol.* 73, 479–491. doi:10.1007/s00248-016-0866-x.

Dukunde, A., Schneider, D., Lu, M., Brady, S., and Daniel, R. (2017). A novel, versatile family IV carboxylesterase exhibits high stability and activity in a broad pH spectrum. *Biotechnol. Lett.* 39, 577–587. doi:10.1007/s10529-016-2282-1.

El-Hofi, M., El-Tanboly, E.-S., and Abd-Rabou, N. S. (2011). Internet Journal of Food Safety Industrial Application of Lipases in Cheese Making: A review. *Internet J. Food Saf.* 13, 293–302.

Facchin, S., Alves, P. D. D., Siqueira, F. de F., Barroca, T. M., Victória, J. M. N., and Kalapothakis, E. (2013). Biodiversity and secretion of enzymes with potential utility in wastewater treatment. *Open J. Ecol.* 3, 34–37. doi:10.4236/oje.2013.31005.

Feller, G., Thiry, M., and Gerday, C. (1990). Sequence of a lipase gene from the antarctic psychrotroph Moraxella TA144. *Nucleic Acids Res.* 18, 6431. doi:10.1093/nar/18.21.6431.

Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., et al. (2016). Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransformation* 2422, 1–15. doi:10.3109/10242422.2016.1151416.

Fojan, P., Jonson, P. H., Petersen, M. T. N., and Petersen, S. B. (2000). What distinguishes an esterase from a lipase: A novel structural approach. *Biochimie* 82, 1033–1041. doi:10.1016/S0300-9084(00)01188-3.

Gu, X., Wang, S., Wang, S., Zhao, L.-X., Cao, M., and Feng, Z. (2015). Identification and Characterization of Two Novel Esterases from a Metagenomic Library. *Food Sci. Technol. Res.* 21, 649–657. doi:10.3136/fstr.21.649.

Guerrand, D. (2017). Lipases industrial applications: focus on food and agroindustries. *Ocl* 24, D403. doi:10.1051/ocl/2017031.

Gupta, R., Gupta, N., and Rathi, P. (2004). Bacterial lipases: An overview of production, purification and biochemical properties. *Appl. Microbiol. Biotechnol.* 64, 763–781. doi:10.1007/s00253-004-1568-8.

Hide, W. a, Chan, L., and Li, W. H. (1992). Structure and evolution of the lipase superfamily. *J. Lipid Res.* 33, 167–178.

Hriscu, M., Chiş, L., Toşa, M., and Irimie, F. D. (2013). pH-Profiling of thermoactive lipases and esterases: Caveats and further notes. *Eur. J. Lipid Sci. Technol.* 115, 571–575.

doi:10.1002/ejlt.201200305.

Hwang, H. T., Qi, F., Yuan, C., Zhao, X., Ramkrishna, D., Liu, D., et al. (2014). Lipase-catalyzed process for biodiesel production: Protein engineering and lipase production. *Biotechnol. Bioeng.* 111, 639–653. doi:10.1002/bit.25162.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi:10.1186/1471-2105-11-119.

Jeon, J. H., Kim, J. T., Kang, S. G., Lee, J. H., and Kim, S. J. (2009). Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar. Biotechnol.* 11, 307–316. doi:10.1007/s10126-008-9145-2.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010.

Khan, F. I., Lan, D., Durrani, R., Huan, W., Zhao, Z., and Wang, Y. (2017). The Lid Domain in Lipases: Structural and Functional Determinant of Enzymatic Properties. *Front. Bioeng. Biotechnol.* 5, 1–13. doi:10.3389/fbioe.2017.00016.

Kohli, P., and Gupta, R. (2016). Medical aspects of esterases: A mini review. *Int. J. Pharm. Pharm. Sci.* 8, 21–26.

Kovacic, F., Mandrysch, A., Poojari, C., Strodel, B., and Jaeger, K. E. (2015). Structural features determining thermal adaptation of esterases. *Protein Eng. Des. Sel.* 29, 65–76. doi:10.1093/protein/gzv061.

Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Eng. Des. Sel.* 13, 179–191. doi:10.1093/protein/13.3.179.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.

Lee, M. H., Khan, R., Tao, W., Choi, K., Lee, S. Y., Lee, J. W., et al. (2018). Soil metagenome-derived 3-hydroxypalmitic acid methyl ester hydrolases suppress extracellular polysaccharide production in Ralstonia solanacearum. *J. Biotechnol.* 270, 30–38. doi:10.1016/j.jbiotec.2018.01.023.

Lee, M. H., Lee, C. H., Oh, T. K., Song, J. K., and Yoon, J. H. (2006). Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: Evidence for a new family of bacterial lipases. *Appl. Environ. Microbiol.* 72, 7406–7409.

doi:10.1128/AEM.01157-06.

Lee, M. H., and Lee, S.-W. (2013). Bioprospecting Potential of the Soil Metagenome: Novel Enzymes and Bioactivities. *Genomics Inform.* 11, 114. doi:10.5808/GI.2013.11.3.114.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., and Chatonnet, A. (2013). ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Res.* 41, 423–429. doi:10.1093/nar/gks1154.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., et al. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50, 437–50. doi:10.1002/prot.10286.

Nardini, M., and Dijkstra, B. W. (1999). α/β hydrolase fold enzymes: The family keeps growing. *Curr. Opin. Struct. Biol.* 9, 732–737. doi:10.1016/S0959-440X(99)00037-8.

Newman, J. R., and Fuqua, C. (1999). Broad-host-range expression vectors that carry the L-arabinose-inducible Escherichia coli araBAD promoter and the araC regulator. *Gene* 227, 197–203. doi:10.1016/S0378-1119(98)00601-5.

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. doi:10.1006/jmbi.2000.4042.

Ovchinnikov, S., Park, H., Varghese, N., Huang, P.-S., Pavlopoulos, G. A., Kim, D. E., et al. (2017). Protein structure determination using metagenome sequence data. *Science (80-.).* 355, 294–298. doi:10.1126/science.aah4043.

Pereira, M. R., Mercaldi, G. F., Maester, T. C., Balan, A., and De Macedo Lemos, E. G. (2015). Est16, a new esterase isolated from a metagenomic library of a microbial consortium specializing in diesel oil degradation. *PLoS One* 10, 1–16. doi:10.1371/journal.pone.0133723.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5. doi:10.1371/journal.pone.0009490.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320-324. doi:10.1093/nar/gku316.

Salis, A., Monduzzi, M., and Solinas, V. (2007). *Industrial Enzymes*. doi:10.1007/1-4020-5377-0.

Sarrouh, B., Santos, T. M., Miyoshi, A., Dias, R., and Azevedo, V. (2012). Up-To-Date Insight on Industrial Enzymes Applications and Global Market. doi:10.4172/2155-9821.S4-002.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* 78, 1606–1619. doi:10.1016/S0006-3495(00)76713-0.

Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* 320, 104–124. doi:10.1016/S0003-2697(03)00289-6.

Stroobants, A., Martin, R., Roosens, L., Portetelle, D., and Vandenbol, M. (2015). New lipolytic enzymes identified by screening two metagenomic libraries derived from the soil of a winter wheat field | Nouvelles enzymes lipolytiques identifiées par criblage de deux banques métagénomiques construites à partir du sol d'un champ de blé d'. *Biotechnol. Agron. Soc. Environ.* 19, 125–131.

Torsvik, V., Goksoyr, J., and Daae, F. L. (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56, 782–787. doi:10.1017/CBO9781107415324.004.

Tyzack, J. D., Furnham, N., Sillitoe, I., Orengo, C. M., and Thornton, J. M. (2017). Understanding enzyme function evolution from a computational perspective. *Curr. Opin. Struct. Biol.* 47, 131–139. doi:10.1016/j.sbi.2017.08.003.

Valero, F. (2012). Lipases and Phospholipases. 861, 161–178. doi:10.1007/978-1-61779-600-5.

Valle, A., Pérez-Socas, L. B., Canet, L., Hervis, Y. D. L. P., De Armas-Guitart, G., Martins-De-Sa, D., et al. (2018). Self-homodimerization of an actinoporin by disulfide bridging reveals implications for their structure and pore formation. *Sci. Rep.* 8, 1–18. doi:10.1038/s41598-018-24688-2.

Webb, B., and Sali, A. (2014). Protein structure modeling with MODELLER. *Methods Mol Biol* 1137, 1–15. doi:10.1007/978-1-4939-0366-5_1.

Xing, M. N., Zhang, X. Z., and Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnol. Adv.* 30, 920–929. doi:10.1016/j.biotechadv.2012.01.021.

Yang, X., Wu, L., Xu, Y., Ke, C., Hu, F., Xiao, X., et al. (2018). Identification and characterization of a novel alkalistable and salt-tolerant esterase from the deep-sea hydrothermal vent of the East Pacific Rise. *Microbiologyopen*, e00601. doi:10.1002/mbo3.601.

Yin, D. L. (Tyler), Bernhardt, P., Morley, K. L., Jiang, Y., Cheeseman, J. D., Purpero, V., et al. (2010). Switching Catalysis from Hydrolysis to Perhydrolysis in *Pseudomonas fluorescens* Esterase ˆ. *Biochemistry* 49, 1931–1942. doi:10.1021/bi9021268.

Zehani, N., Dzyadevych, S. V., Kherrat, R., and Jaffrezic-Renault, N. J. (2014). Sensitive impedimetric biosensor for direct detection of diazinon based on lipases. *Front. Chem.* 2, 1–7. doi:10.3389/fchem.2014.00044.