

**MARIANA RANGEL PEREIRA**

**CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ENZIMAS  
LIPOLÍTICAS DE UM CONSÓRCIO MICROBIANO DEGRADADOR  
DE ÓLEO DIESEL**

Tese apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia USP/Instituto Butantan/IPT, para obtenção do Título de Doutor em Biotecnologia.

São Paulo  
2015

**MARIANA RANGEL PEREIRA**

**CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ENZIMAS  
LIPOLÍTICAS DE UM CONSÓRCIO MICROBIANO DEGRADADOR  
DE ÓLEO DIESEL**

Tese apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia USP/Instituto Butantan/IPT, para obtenção do Título de Doutor em Biotecnologia.

Área de concentração: Biotecnologia

Orientadora:

Dra. Eliana Gertrudes de Macedo Lemos

Co-orientadora:

Dra. Andrea Balan Fernandes

Versão original.

São Paulo  
2015

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
Serviço de Biblioteca e Informação Biomédica do  
Instituto de Ciências Biomédicas da Universidade de São Paulo

reprodução não autorizada pelo autor

Pereira, Mariana Rangel.

Caracterização funcional e estrutural de enzimas lipolíticas de um consórcio microbiano degradador de óleo diesel / Mariana Rangel Pereira. -- São Paulo, 2015.

Orientador: Profa. Dra. Eliana Gertrudes de Macedo Lemos.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan. Área de concentração: Biotecnologia. Linha de pesquisa: Bioquímica de Microorganismos.

Versão do título para o inglês: Functional and structural characterization of lipolytic enzymes from a microbe consortium specialized for diesel oil degradation.

1. Metagenoma 2. Consórcio microbiano 3. Esterase 4. Lipase I. Lemos, Profa. Dra. Eliana Gertrudes de Macedo II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan III. Título.

ICB/SBIB019/2015

UNIVERSIDADE DE SÃO PAULO  
**Programa de Pós-Graduação Interunidades em Biotecnologia**  
Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

---

Candidato(a): Mariana Rangel Pereira.

Título da Tese: Caracterização funcional e estrutural de enzimas lipolíticas de um consórcio microbiano degradador de óleo diesel.

Orientador(a): Profa. Dra. Eliana Gertrudes de Macedo Lemos.

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a ...../...../....., considerou

**Aprovado(a)**

**Reprovado(a)**

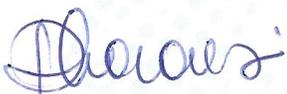
Examinador(a): Assinatura: .....  
Nome: .....  
Instituição: .....

Presidente: Assinatura: .....  
Nome: .....  
Instituição: .....

**CEUA- COMISSÃO DE ÉTICA NO USO DE ANIMAIS****CERTIFICADO**

Certificamos que o Protocolo n° 022047/14 do trabalho de pesquisa intitulado “Caracterização funcional e estrutural de enzimas lipolíticas de um consórcio microbiano degradador de óleo diesel”, sob a responsabilidade da Prof<sup>a</sup> Dr<sup>a</sup> Eliana Gertrudes de Macedo Lemos, foi analisado pela COMISSÃO DE ÉTICA NO USO DE ANIMAIS (CEUA), em reunião no dia 07 de novembro de 2014. O projeto não se enquadra nas responsabilidades da CEUA, visto que não utiliza animais no seu escopo.

Jaboticabal, 07 de novembro de 2014

---

Prof<sup>a</sup> Dra Paola Castro Moraes  
Coordenadora CEUA

Comissão de Ética no uso de Animais  
CEUA/CNPEM

*Certificado de Isenção*

Certificamos que o projeto **“Caracterização Funcional e Estrutural de Enzimas Lipolíticas de um Consórcio Microbiano Degradador de Óleo Diesel”** (protocolo nº **04**), sob responsabilidade de **Andrea Balan Fernandes**, foi analisado na presente data pela CEUA- Comissão de Ética no uso de Animais, tendo sido deliberado que o referido projeto não envolve manipulação animal que justifique a aprovação quanto aos princípios éticos exigidos pela Comissão.

Campinas, 21 de maio de 2013



Dra. Carolina F. M. Z. Clemente  
Presidente



Dra. Maria Carolina Scatolin do Rio  
Bióloga

**Formulário de encaminhamento de projetos de pesquisa para análise pela CIBio - Comissão Interna de Biossegurança da ABTLuS – Associação Brasileira de Tecnologia de Luz Síncrotron**

Pesquisador responsável: Caracterização funcional e estrutural de enzimas lipolíticas de um consórcio microbiano degradador de óleo diesel

Pesquisador responsável: Andrea Balan

Experimentador: Mariana Rangel Pereira

Nível do treinamento do experimentador: [ ]-Iniciação científica, [ ]-mestrado, [ X ]-doutorado, [ ]-doutorado direto, [ ]-pós-doutorado, [ ]-nível técnico, [ ]-outro, especifique: \_\_\_\_\_

Resumo do projeto:

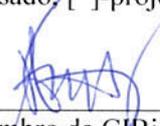
A utilização de micro-organismos para o processamento de materiais naturais é realizada há décadas e vem crescendo com as novas descobertas em termos de métodos de isolamento, caracterização e cultivo destes. Estudos de populações microbianas de diferentes habitats, utilizando seqüências conservadas dos genes *16S rRNA* e *18S rRNA* mostra que a maior parte dos micro-organismos que compõem a biosfera ainda não foi estudada. Esta diversidade populacional e genética vem sendo objeto de estudo em diferentes aspectos, mas principalmente com o objetivo de identificar microrganismos com potencial de biorremediação, farmacêutico e biotecnológico. Uma das estratégias mais empregadas que oferece uma combinação quase ilimitada para encontrar novos genes codificadores de produtos relevantes é a abordagem metagenômica, a partir da qual o DNA genômico de microrganismos presentes em amostras do ambiente é isolado e clonado em vetores específicos para a construção de bibliotecas.

Em função do enorme potencial biotecnológico, as enzimas lipolíticas vêm atraindo atenção no mercado global, como por exemplo: nas misturas enzimáticas para formulação de detergentes; na indústria de couro, podendo ser aplicada juntamente com outras hidrolases na remoção de pêlos e gorduras subcutâneas; na produção de cosméticos, fármacos, aromas, biodiesel, dentre outros. As lipases microbianas apresentam baixa toxicidade, são facilmente biodegradáveis e trabalham em condições amenas, destacando-se pela quimio-seletividade, o que reduz os efeitos colaterais dos fármacos. Com o objetivo de se identificar enzimas com potencial biotecnológico, amostras de um consórcio bacteriano degradador de óleo diesel, provenientes de uma região bastante afetada por resíduos químicos, foram usadas como material para a bioprospecção, usando a abordagem metagenômica (projeto de mestrado da aluna, processo número 2009/06991-0). A partir da análise dos quadros de leitura, foram identificadas quatro enzimas com função de lipase que apresentam baixa similaridade com seqüências de ortólogos depositadas nos principais bancos de seqüências conhecidos. A caracterização preliminar destas enzimas revelou seu alto potencial biotecnológico e neste contexto, o presente projeto tem por objetivo produzir em larga escala, caracterizar bioquimicamente e cristalizar tais enzimas lipolíticas.

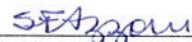
A CIBio analisou este projeto em reunião realizada no dia: 9.1.12.

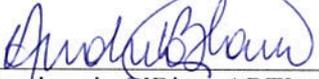
Parecer final: [ X ]-projeto aprovado, [ ]-projeto recusado, [ ]-projeto com deficiências, favor comentários anexo.

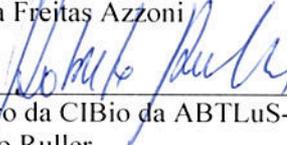
  
Presidente de CIBio – ABTLuS-LNBio  
Jörg Kobarg

  
Membro da CIBio da ABTLuS-CTBE  
Fabio Squina

  
Membro da CIBio – ABTLuS-LNBio  
Prof. Dr. Celso Eduardo Benedetti

  
Membro da CIBio da ABTLuS-CTBE  
Sindelia Freitas Azzoni

  
Membro da CIBio – ABTLuS-LNBio  
Andrea Balan

  
Membro da CIBio da ABTLuS-CTBE  
Roberto Ruller

  
Membro da CIBio – ABTLuS-SGT  
Carolina Barbosa Marini

## ACKNOWLEDGMENTS

I extend my thanks to Professor Doctor Eliana Gertrudes de Macedo Lemos and Professor Doctor Andrea Balan Fernandes for their excellent guidance and supervision in this work;

To Dr. Marko Hyvonen for hosting me in his laboratory and for greatly contributing to the structural characterization of the proteins in this study;

To Gustavo Fernando Mercaldi and Thaís Carvalho Maester for their contributions to experiments;

To the members of the Biochemistry Laboratory of Microorganisms and Plants (LBMP), the Laboratory of Applied Structural Biology (LBEA) and Marko Hyvonen's group for their attention, friendship and scientific knowledge;

To the Department of Technology, Jaboticabal Campus, Brazil; to the National Laboratory of Biosciences (LNBio), Campinas, Brazil; and the Department of Biochemistry, University of Cambridge, England, for the facilities used in the execution of experiments;

To the technicians, staff and academics of the involved institutions who helped in the running of this Project for always being solicitous and helpful;

To the São Paulo Research Foundation (FAPESP) for funding and the awarded bursary.

This project was supported by the São Paulo Research Foundation (FAPESP), process number 2011/09136-7, and it was developed at Biochemistry Laboratory of Microorganisms and Plants (LBMP), Department of Technology, Jaboticabal Campus, São Paulo State University (UNESP); and at National Laboratory of Biosciences (LNBio), Campinas, Brazil. In the same way, FAPESP supported the Research Internships Abroad (BEPE), process number 2012/20490-0, which was conducted at University of Cambridge, England.

*“ We ourselves feel that what we are doing is just a drop in the ocean.  
But the ocean would be less because of that missing drop.”*

*Mother Teresa of Calcutta*

## RESUMO

PEREIRA, M. R. **Caracterização funcional e estrutural de enzimas lipolíticas de um consórcio microbiano degradador de óleo diesel**. 2015. 115 f. Tese (Doutorado em Biotecnologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2015.

O comércio mundial de enzimas industriais estava estimado em 2.3 bilhões de dólares entre detergentes (US\$ 789 milhões), aplicações alimentícias (US\$ 634 milhões), agricultura (US\$ 237 milhões), entre outros. Neste contexto, as enzimas lipolíticas estão atraindo enorme atenção devido ao seu potencial biotecnológico, visto que estas podem catalisar múltiplas reações (hidrólise, acidólise, interesterificação e glicerólise). Enzimas lipolíticas de origem microbiana são economicamente atrativas por serem biodegradáveis, atuarem normalmente em condições brandas, e serem quimio-seletivas propiciando à indústria farmacêutica a obtenção de drogas com efeito colateral reduzido. Em trabalho anterior (dissertação de mestrado, número de processo 2009/06991-0), cinco genes potenciais codificadores de esterases/lipases foram identificados através de uma biblioteca metagenômica obtida de um consórcio microbiano degradador de óleo diesel. Em continuação, neste projeto, quatro destes genes foram clonados em vetores de expressão e expressos em *Escherichia coli* BL21 (DE3) a partir dos vetores pET28a ou pHAT2, e as proteínas correspondentes, denominadas como Est8, Est16, *ORF17* e Est30, foram submetidas a ensaios funcionais e estruturais. Nossos resultados mostram que todas as proteínas puderam ser expressas e purificadas na forma solúvel e estável, sendo que a caracterização funcional da Est8, Est16 e Est30 mostrou que estas proteínas são esterases que possuem alta eficiência catalítica para *p*-nitrofenil acetato, *p*-nitrofenil butirato e *p*-nitrofenil valerato, respectivamente. Interessantemente, estas proteínas apresentaram elevada atividade em pH 9.0 e temperatura ótima entre 50-60 °C, em concordância com os ensaios de desnaturação térmica. Ressalta-se que a atividade da proteína Est16 aumentou na presença de dimetil sulfoxido (DMSO) e a da proteína Est8 não foi afetada por concentrações de até 10% deste solvente, tornando-as atrativas para serem usadas com solventes orgânicos e em elevadas temperaturas. Ainda, três estruturas cristalográficas foram resolvidas (Est8 a 1.8 Å, *ORF17* a 1.4 Å; e Est30 a 1.7 Å e em complexo com butirato a 1.65 Å), o que permitiu a caracterização da cavidade interna destas proteínas, dos resíduos envolvidos na ligação com o substrato e o posicionamento das tríades catalíticas. Estes dados possibilitaram a construção de um conjunto de mutantes: o mutante Est8MF, onde se visou aumentar o canal que dá acesso à tríade catalítica da proteína Est8; e os mutantes Est30asp19 e Est30ala34, que foram construídos a fim de expor a tríade catalítica da proteína Est30. Os ensaios funcionais do mutante Est8MF mostraram que a sua velocidade de reação foi quatro vezes maior quando comparada com a proteína selvagem (Est8) usando como substrato o *p*-nitrofenil butirato, entretanto, para *p*-nitrofenil acetato a velocidade foi reduzida. Adicionalmente, as estruturas cristalográficas resolvidas neste trabalho, quando comparadas com as estruturas de esterase/lipase depositadas no Protein Data Bank (PDB), revelaram interessantes características para futuras aplicações biotecnológicas.

**Palavras-chave:** Metagenoma. Consórcio microbiano. Esterase. Lipase.

## ABSTRACT

PEREIRA, M. R. **Functional and structural characterization of lipolytic enzymes from a microbe consortium specialized for diesel oil degradation.** 2015. 115 f. Thesis (Ph.D. in Biotechnology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2015.

The global trade of industrial enzymes is estimated at 2.3 billion U.S. dollars, divided mainly between detergents (US\$ 789 million), food applications (US\$ 634 million), and agriculture (US\$ 237 million). Within this trade, lipolytic enzymes have attracted enormous attention because of their biotechnological potential as catalysts of multiple reaction types (including hydrolysis, acidolysis, interesterification and glycerolysis). Lipolytic enzymes of microbial origin are economically attractive because they are easily biodegradable, usually act in mild conditions, and are chemo-selective, providing the pharmaceutical industry a method for obtaining drugs with reduced side effects. In a previous work (master's student project, process number 2009/06991-0), five genes encoding putative esterases/lipases were identified in a metagenomic library obtained from a microbe consortium isolated from diesel oil-contaminated soil. Here, four individual genes were cloned into the pET28a or pHAT2 vector and expressed in *Escherichia coli* BL21 (DE3), and their corresponding recombinant proteins, designated Est8, Est16, ORF17 and Est30, were used for functional and structural studies. Our results showed that all proteins could be expressed and purified as stable and soluble enzymes. Through functional characterization, it was possible to verify that Est8, Est16 and Est30 are esterases with the highest catalytic efficiency against *p*-nitrophenyl acetate, *p*-nitrophenyl butyrate and *p*-nitrophenyl valerate, respectively. Interestingly, they displayed the highest activity at pH 9.0 and their optimal temperatures were approximately 50-60 °C, which are in agreement with the results of the thermal denaturation assay. In addition, Est16 had enhanced activity in the presence of dimethyl sulfoxide (DMSO), while Est8 showed no difference in activity in the presence of up to 10% DMSO. Thus, these thermostable enzymes are very attractive for use with organic solvents and at elevated temperatures. Crystal structures of the three proteins were also obtained (Est8 at 1.8 Å; ORF17 at 1.4 Å; and Est30 in a free form at 1.7 Å and in a complex with butyrate at 1.65 Å). The structures allowed for the characterization of the pocket of each esterase/lipase, the residues involved in the substrate binding, and the positioning of the catalytic triad. With this information, a set of mutants of these proteins were designed for the first time: in the Est8 protein, the channel for the passage of the substrate was increased in generating the Es8MF mutant, while in the Est30 protein, the catalytic triad was exposed in creating the Est30asp19 and Est30ala34 mutants. The reaction velocity of the Est8MF mutant was shown to be four times faster than that of the wild type (Est8) against *p*-nitrophenyl butyrate. However, it was noticed that Est8MF showed reduced activity toward *p*-nitrophenyl acetate compared to Est8. Interestingly, the comparison between the newly solved structures and the deposited esterase/lipase structures from the Protein Data Bank (PDB) revealed interesting features of the proteins for future biotechnological studies.

**Keywords:** Metagenome. Microbe consortium. Esterase. Lipase.

## LIST OF ABBREVIATIONS

BSA: bovine serum albumin  
CHES: N-cyclohexyl-2-aminoethanesulfonic acid  
DMF: dimethylformamide  
DMSO: dimethyl sulfoxide  
DNA: deoxyribonucleic acid  
EDTA: ethylenediaminetetraacetic acid  
HEPES: 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid  
IMAC: metal ion affinity chromatography  
IPTG: isopropyl  $\beta$ -D-1-thiogalactopyranoside  
OD: optical density  
ORF: open read frame  
PCR: polymerase chain reaction  
*p*-NP: para-nitrophenol  
p/v: weight per volume  
pH:  $\log [H^+]$   
r.s.m.d: root-mean-square deviation  
SDS: sodium dodecyl sulfate  
SEC: size-exclusion chromatography  
Tris: Tris(hydroxymethyl)aminomethane  
UV: ultraviolet light  
v/v: volume per volume

## LIST OF UNITS

Å: angstrom  
°C: degree Celsius  
g: acceleration of gravity  
g: gram  
kb: kilobase  
kDa: kilodalton  
L: liter  
M: molar  
Mb: megabase pair  
mg: milligram  
mL: milliliter  
mM: millimolar  
ng: nanogram  
nm: nanometer  
nM: nanomolar  
pb: base pair  
µg: microgram  
µL: microliter  
µM: micromolar

## SUMMARY

<b>1 INTRODUCTION</b> .....	17
<b>2 AIMS</b> .....	22
<b>3 MATERIALS AND METHODS</b> .....	23
<b>3.1 Search for patents and enzyme global market data</b> .....	23
<b>3.2 Cloning of <i>orf17</i> and <i>est30</i> genes</b> .....	24
<b>3.3 Construction of mutants</b> .....	25
<b>3.4 Expression tests</b> .....	26
<b>3.5 Overexpression and purification</b> .....	26
<b>3.6 Circular dichroism of Est30</b> .....	27
<b>3.7 pH influence on the Est30 protein in the tributyrin assay</b> .....	27
<b>3.8 Functional characterization of Est30</b> .....	28
<b>3.9 Protein crystallization and data collection</b> .....	29
<b>3.10 Data processing and structure solving and refinement</b> .....	30
<b>4 RESULTS AND DISCUSSION</b> .....	32
<b>4.1 Chapter 1: Evaluation of the global market and the search for patented sequences of esterases and lipases</b> .....	32
<b>4.2 Chapter 2: Est16 protein</b> .....	36
<b>4.2.1 Abstract</b> .....	36
<b>4.2.2 Introduction</b> .....	37
<b>4.2.3 Methods</b> .....	38
<b>4.2.3.1 Sampling</b> .....	38
<b>4.2.3.2 Screening for lipolytic activity</b> .....	39
<b>4.2.3.3 Selection of the clone and sub-library construction</b> .....	39
<b>4.2.3.4 DNA sequencing, gene annotation and qualitative analysis of the Est16 activity</b> .....	39
<b>4.2.3.5 Phylogenetic analysis</b> .....	40
<b>4.2.3.6 Molecular modeling of Est16 and structural analysis</b> .....	40
<b>4.2.3.7 Cloning, expression and purification of Est16</b> .....	41
<b>4.2.3.8 Tributyrin assay in Petri dishes</b> .....	42

4.2.3.9 Measuring the enzymatic characteristics of Est16.....	42
4.2.3.10 Circular dichroism .....	43
4.2.3.11 Nucleotide sequence accession number .....	44
<b>4.2.4 Results and Discussion</b> .....	44
4.2.4.1 Metagenomic library construction and screening for lipolytic activity .....	44
4.2.4.2 Est16 shares sequence and structure similarities with members of family V and with the patented Anti-Kazlauskas lipase .....	46
4.2.4.3 The structural features of Est16 model corroborates the sequence similarities with <i>Pseudomonas fluorescens</i> and the Anti-Kazlauskas lipase .....	48
4.2.4.4 Est16 was expressed and purified as a soluble, monodisperse and stable protein.....	49
4.2.4.5 Purified Est16 is active in basic pH and thermostable .....	50
4.2.4.6 Substrate specificity and kinetic parameters of Est16.....	51
<b>4.2.5 Conclusions</b> .....	54
<b>4.2.6 Acknowledgments</b> .....	54
<b>4.2.7 References</b> .....	54
<b>4.2.8 Supplementary material</b> .....	58
<b>4.3 Chapter 3: Est8 protein</b> .....	61
<b>4.3.1 Abstract</b> .....	62
<b>4.3.2 Introduction</b> .....	63
<b>4.3.3 Materials and methods</b> .....	64
4.3.3.1 Screening for lipolytic activity.....	64
4.3.3.2 Sequence analysis of est8 gene .....	64
4.3.3.3 DNA amplification and plasmid construction .....	65
4.3.3.4 Protein expression, extraction and purification .....	65
4.3.3.5 Circular dichroism and spectroscopic analysis .....	66
4.3.3.6 Enzymatic characterization of Est8.....	67
4.3.3.7 Est8 crystallization .....	68
4.3.3.8 Data collection, processing and structure refinement.....	68
<b>4.3.4 Results and discussion</b> .....	68
4.3.4.1 Est8 is a family IV member of esterases and lipases .....	68

4.3.4.2 Recombinant Est8 was expressed as a soluble, folded and stable protein.....	71
4.3.4.3 Kinetics characterization of Est8 .....	73
4.3.4.4 Crystallization and structure determination .....	76
4.3.4.5 The opening of the channel for substrate in Est8 increased its activity against C4 and C5.....	79
<b>4.3.5 Conclusion .....</b>	<b>82</b>
<b>4.3.6 Acknowledgments .....</b>	<b>82</b>
<b>4.3.7 References .....</b>	<b>82</b>
<b>4.3.8 Supplementary material .....</b>	<b>86</b>
<b>4.4 Chapter 4: Est30 protein.....</b>	<b>88</b>
4.4.1 Expression and purification .....	88
4.4.2 Assay in Petri dishes containing tributyrin and circular dichroism.....	90
4.4.3 Substrate specificity and kinetic parameters of Est30.....	91
4.4.4 Crystallization trials and diffraction .....	92
4.4.5 Construction of Est30asp19 and Est30ala34 mutants .....	98
<b>4.5 Chapter 5: ORF17 protein.....</b>	<b>100</b>
4.5.1 Expression and purification .....	100
4.5.2 Crystallization .....	102
4.5.3 Three-dimensional structure of ORF17 .....	103
<b>4.6 Chapter 6: Structural comparison and perspectives.....</b>	<b>105</b>
<b>5 CONCLUSIONS.....</b>	<b>107</b>
<b>RERENCES .....</b>	<b>108</b>
<b>Annex A - List of patented esterases/lipases used in building the phylogenetic tree containing the four enzymes studied in this Ph.D. project.....</b>	<b>113</b>
<b>Annex B - The co-authors' authorization to use the Est16 manuscript as a chapter in this Ph.D. thesis. ....</b>	<b>114</b>
<b>Annex C - The co-authors' authorization to use the Est8 manuscript as a chapter in this Ph.D. thesis. ....</b>	<b>115</b>

## 1 INTRODUCTION

The biosphere is dominated by microorganisms that are still mostly unknown. It is estimated that 99% of prokaryotic organisms have not been studied, which is probably due to the traditional methods of cultivation, which limit the development of these organisms in laboratory conditions (HARDEMAN; SJÖLING, 2007; HENNE et al., 2000; ROH; VILLATTE, 2008). The recent advances in molecular microbial ecology research have provided strong evidence of the existence of new microorganism species in the environment in large numbers and with high diversity (STEELE; STREIT, 2005), especially when compared to those few species that are cultivable in the laboratory. Confirmation of this evidence comes from the estimation of the DNA complexity and the discovery of several unique sequences of *16S rRNA* and *18S rRNA* from different environmental sources (RONDON et al., 2000; SILVEIRA et al., 2006).

To access the genetic resources of the vast numbers of microbial species that have thus far escaped detailed scientific testing, researchers have started using a new approach termed metagenomics (HANDELSMAN et al., 1998). This strategy includes the genomic DNA extraction from environmental samples, known as the collective genome, followed by the cloning of the obtained material (RONDON et al., 2000). The results are libraries composed of small or large DNA fragments which can be cloned into appropriate vectors to produce clones that can be used to examine metabolic routes, to analyze microbial diversity, to identify genes encoding proteins of biotechnological interest, and other purposes.

Previous studies have shown that these metagenomic analyses offer unlimited combinations for identifying new genes encoding potential enzymes such as esterases and lipases (COUTO et al., 2010; HU et al., 2010; WU; SUN, 2009). For example, in the last few years, several genes encoding lipolytic enzymes were identified from metagenomic libraries that were taken from various environmental samples, including soil (CHOW et al., 2012; LEE et al., 2004), soil collected from a wastewater treatment plant (GLOGAUER et al., 2011), mangrove sediment (COUTO et al., 2010), marine sediment (HU et al., 2010; JEON et al., 2012; PENG et al., 2014), swamp sediment (SEO et al., 2014), pond and lake water (RANJAN et al., 2005), sea water (CHU et al., 2008; FANG et al., 2014), and river water (WU; SUN, 2009).

Among these discoveries, Fang and collaborators (2014) identified the Est9x esterase from a marine microbial metagenome. This enzyme is a member of a new family of microbial lipolytic enzymes proposed by these authors that presents a high activity against *p*-nitrophenyl

acetate, an optimal temperature at 65 °C, and high salt tolerance, all of which make it an enzyme with potential for industrial applications. Another esterase termed Est23 was identified in a soil metagenomic library and demonstrated stability in the presence of up to 50% benzene or alkanes (JIN et al., 2012). In this context, the Est2K esterase showed activity in the presence of 30% methanol and was mostly active at pH 10.0, suggesting its potential as an alkaline enzyme (KIM et al., 2010).

In terms of the enzyme market, it is known that the global enzyme trade was estimated at 2.3 billion dollars, whose the main profit was divided into detergents (US\$ 789 million), food applications (US\$ 634 million), agriculture (US\$ 237 million), among others (Global Industry Analysts, 2004). The current enzyme market is estimated at 3.4 billion euros, with an annual growth of 6.5 to 10% (CBDM.T, 2008). Within this market, lipolytic enzymes have attracted significant attention due to their biotechnological potential (VAKHLU; KOUR, 2006).

Most of the lipolytic enzymes used in industry are of microbial origins (JAEGER; DIJKSTRA; REETZ, 1999). The microorganism-derived enzymes are available for industrial application and the synthesis of new products, enabling high gains in different biotechnology processes. These enzymes have greater efficiency reducing expenses, and use fewer caustic chemicals and thus are beneficial to the environment (LORENZ; ECK, 2005). Nevertheless, the largest problem facing the use of lipases in industrial processes is the costly and time-consuming purification process, which generally result in a loss of enzymatic activity. However, this situation has the potential to change due to recent advances in technology, such as genetic engineering as well as protein modification and immobilization (KOBBLITZ; PASTORE, 2004; SAXENA et al. 2003).

Based on their substrate preferences, these lipolytic enzymes are categorized as esterases (EC 3.1.1.1) and lipases (EC 3.1.1.3), which catalyze the hydrolysis and synthesis of short (carbon chain length  $\leq 10$ ) and long-chain acylglycerols (carbon chain length  $> 10$ ), respectively. In addition, the bacterial lipolytic enzymes can be classified based on their amino acid sequences and some fundamental biological properties. According to Arpigny and Jaeger (1999), these enzymes are classified into eight major families: family I (true lipases); family II, which has members with the conserved GDSL motif; family III, which has representatives showing a 20% identity with the human PAF-AH protein; family IV, known as the hormone-sensitive lipases (HSL); family V, which has mesophilic, cold or heat-adapted representatives; family VI, consisting of enzymes that have a 40% sequence similarity with the eukaryotic lysophospholipases and presents the smallest esterases known (23-26 kDa);

family VII, which has enzymes with sequence similarity to eukaryotic acetylcholine esterases and carboxylesterases of approximately 30-40%; and family VIII, which has enzymes that present the conserved motif Ser-X-X-Gly (where X can be any amino acid) and enzymes displaying similarities to several class C  $\beta$ -lactamases. Despite this proposed classification, new lipases and esterases have been identified from different environmental samples using the metagenomic approach. According to Ranjan and collaborators (2005), the incessant search for esterases and lipases will increase the diversity of known lipolytic enzymes and consequently the number of families.

Esterases and lipases are members of the family of enzymes containing an alpha-beta hydrolase fold. This family includes proteases, dehalogenases, epoxide hydrolases and peroxidases, making it the most versatile enzyme family which contains the most common protein fold. Enzymes in this family do not share sequence similarity and substrate preferences, however, their structural arrangements are relatively homologous and they share a preserved catalytic site, suggesting a possible common ancestor (OLLIS et al., 1992). The folding of an alpha-beta hydrolase protein is characterized by a core of eight  $\beta$ -sheets connected by  $\alpha$ -helices, which gives the arrangement  $\alpha/\beta/\alpha$ . In most members of this family, the  $\beta$ -sheets are parallel, but some proteins show reverse orientations of the  $\beta$ -sheets that result in an antiparallel orientation. This folding can provide a stable framework.

The catalytic residues are always composed of a highly conserved Ser-Asp-His triad. The nucleophilic residue is positioned after the  $\beta$ 5-strand, the acid residue (Glu instead of Asp in some enzymes) is often positioned after the  $\beta$ 7-strand, and the absolutely conserved histidine residue is located after the last  $\beta$ -strand (NARDINI; DIJSTRA, 1999), however, the location of the histidine residue (sequence length) may differ considerably between the different members in this family.

The nucleophilic residue is always located on a sharp bend called a "nucleophilic elbow", which facilitates the interaction with the substrate. The geometry of the "nucleophilic elbow" contributes to the formation of the oxy-anion binding site, which is necessary for the stabilization of the transition state during hydrolysis

Esterases can be differentiated from lipases mainly due to their substrate specificity and the lack of an interfacial activation (LONG, 1971), which is observed only in lipases. Lipases present a hydrophobic domain (lid) that covers the catalytic triad and moves to expose the active site in a minimal concentration of substrate (BORNSCHEUER, 2002). In contrast, esterases follow Michaelis-Menten kinetics. However, esterases and lipases do not differ in the mechanisms of ester hydrolysis and synthesis, which occur in a few steps. When

the substrate is bound to the serine residue, a tetrahedral intermediate is produced that is stabilized by the His and Asp residues. An acyl enzyme complex is formed after the release of alcohol, and then the nucleophile residue is attacked by water, alcohol, or an ester depending on the reaction (hydrolysis, transesterification or esterification, respectively). Again, the result is a tetrahedral intermediate that gives rise to the final product, which is either an acid or an ester, and the free enzyme (BORNSCHEUER, 2002).

Esterases and lipases present a broad applicability in different sectors, which justifies the incessant search for and research on these enzymes. In last few years, several esterases have become available from different companies, such as Amano, Diversa and Roche Diagnostics. Esterases can be used in the production of pure compounds, in the processing of food and beverages, as a flavor enhancer, in agricultural areas to detoxify insecticide compounds, in paper manufacturing, in cosmetics and perfumes, and in other areas. One such application is the use of a carboxyl esterase (32 kDa) from *Bacillus subtilis* in the development of Naproxen, an anti-inflammatory drug, which was carried out due to the high stereospecificity of S-naproxen [2-(6-methoxy-2-naphthyl)-propionic acid] esters and esters of related drugs (QUAX; BROEKHUIZEN, 1994).

In the food industry, a feruloyl esterase has been used to produce the pentyferulate ester, which is a precursor in food processing as well as in cosmetics (GIULIANI et al., 2001). In the production of sake, an esterase from *Saccharomyces cerevisiae* has been used for the production of isoamyl acetate, which determines the drink flavor (FUKUDA et al. 1998). Moreover, esterases and lipases from *Lactobacillus* sp. have been used to enhance the flavor of cheeses through the hydrolysis of milk fat (CHOI; LEE, 2001). Another important application is the use of phosphotriesterases from *Brevundimonas diminuta* and *Alteromonas* sp. to degrade organophosphorous compounds used in insecticides and nematicides, which have negative environmental impacts (HORNE et al. 2002). Finally, one of the most important applications for lipases is the production of detergents. Companies such as DuPont Industrial Biosciences have successfully developed detergents using bacterial lipases, including Lumafast™ from *Pseudomonas mendocina* and Lipomax™ from *Pseudomonas alcaligenes* (JAEGER; REETZ, 1998). These detergents have shown increased stability against detergent composition in the presence of boronic acid derivatives due to a reversible complex formation in the active site (patent number EP0478050A1).

To explore genes that encode proteins with biotechnological potential, a metagenomic library from a microbe consortia specializing in diesel oil degradation was used. This library is part of a library collection of different environmental sources and has been stored at the

Biochemistry Laboratory of Microorganisms and Plants (LBMP), Department of Technology, at the State University of São Paulo (UNESP), Jaboticabal Campus.

The metagenomic library from the microbial consortia contains approximately 4224 clones and has been used to prospect several genes, including genes that encode lipolytic enzymes. The soil that was used to develop the microbial consortia was obtained from an old lubricant factory in Ribeirão Preto town, São Paulo State, Brazil, which had been disposing of waste into the soil for a period of 15 years.

In the master's project entitled "Screening for lipolytic enzyme codification genes in a metagenomic library of consortia specializing in diesel oil degradation" (process number 2009/06991-0), five genes encoding lipolytic enzymes were identified. In the current Ph.D. project, four of these genes were cloned individually into the expression vector and the recombinant proteins were expressed and purified in order to characterize their functions and structures. One aim of this project was to evaluate the following for each enzyme: the substrate preference; the optimal pH and temperature; the enzyme performance in the presence of additives; the tolerance to organic solvents and detergents; the thermal stability; and the biotechnological potential. The second aim of this project was to design mutants in order to improve the enzymatic activity whenever necessary. The overall goal of this project was to obtain an enzyme collection that could be used in different industrial sectors, including bioremediation.

## 2 AIMS

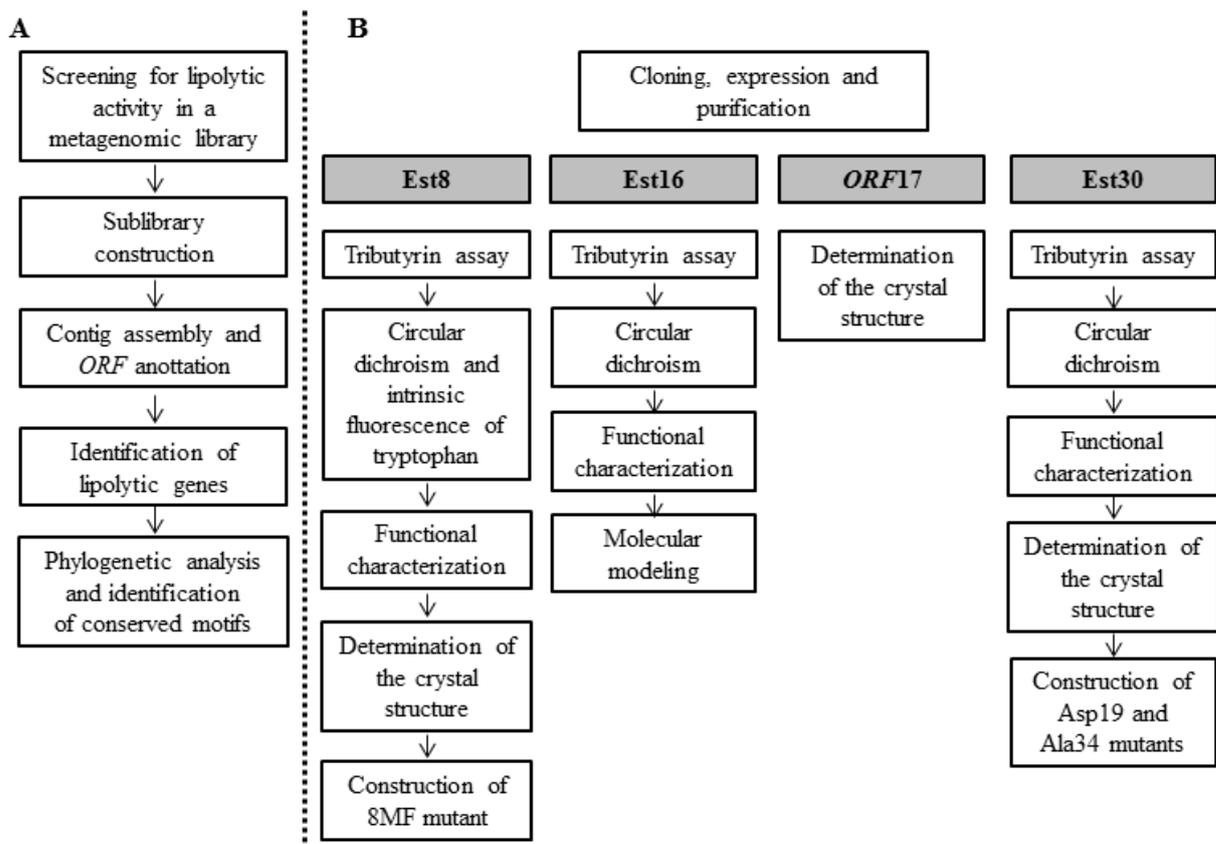
The main objective of this project was to develop the functional and structural characterization of four *ORFs* encoding lipolytic enzymes from a metagenomic microbe consortium specializing in diesel oil degradation.

To meet this goal, the following specific objectives were addressed:

- To clone the genes encoding the enzymes of interest in expression vectors for overexpression and purification;
- To purify the proteins through immobilized metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC);
- To perform spectroscopic analysis of the proteins using circular dichroism and the intrinsic fluorescence of tryptophans;
- To use kinetic analyses to functionally characterize the enzymes;
- To set up crystallization trials for all of the enzymes;
- To determine the three-dimensional structures of the crystallized proteins in the presence and absence of substrates;
- To perform analyses of the structures in order to characterize the substrate-binding pockets and make rational designs for mutations whenever necessary;
- To compare the results of the different proteins in order to determine each of their biotechnological potentials.

### 3 MATERIALS AND METHODS

A pipeline of what was developed in this project is presented as a methodological flowchart including results from Est8, Est16, *ORF17* and Est30 proteins (Figure 1). In addition, detailed material and methods used for obtaining and characterization of Est30 and *ORF17* proteins is shown. On the other hand, the methodology applied for Est16 and Est8 proteins is shown as manuscript format in the chapters 4.2 and 4.3, respectively.



**Figure 1. Methodological flowchart.** (A) Steps that were performed during a master's project (process number 2009/06991-0) to identify lipolytic genes. (B) The activities developed in this work were conducted with the Est8, Est16, *ORF17* and Est30 proteins.

#### 3.1 Search for patents and enzyme global market data

To evaluate the global enzyme market and to search for patented esterases and lipases used in different applications, we used the Questel Intellectual Property Portal (<http://www.orbit.com/#WelcomePage>). The first step was to detect the countries and companies that have been working on esterases and lipases over the last few years. For this, a search command was applied in the Questel Portal: (lipase OR esterase) AND (alpha beta

hydrolase OR amino acid sequence OR hydrolase fold). After this analysis, two search strategies were adopted:

1. The GenomeQuest (<http://www.genomequest.com/>) database was searched for esterases and lipases as previously mentioned, which have also been published in manuscripts or books using their patent code;
2. The Orbit program was searched for patented esterases and lipases using the specific keywords:
  - (lipase or esterase) AND (hydrolase and enzymatic activity and food);
  - (lipase or esterase) AND (hydrolase and enzymatic activity and dairy);
  - (lipase or esterase) AND (hydrolase and enzymatic activity and baked products);
  - (lipase or esterase) AND (hydrolase enzymatic activity and detergents);
  - (lipase or esterase) AND (hydrolase enzymatic activity and fine chemicals);
  - (lipase or esterase) AND (hydrolase and enzymatic activity and pulp);
  - (lipase or esterase) AND (bioremediation);
  - (lipase or esterase) AND (waste treatment);
  - (lipase or esterase) AND (biofuel and bioenergy).

To refine the Questel results, the specific keywords were required to appear in the title, abstract, object of invention and object of study. The patents were selected according to the citation number and/or how many words of interest were contained in their description. The sequences were obtained from the GenomeQuest database using their patent code.

### **3.2 Cloning of *orf17* and *est30* genes**

The two genes (*orf17* and *est30*) were amplified by Polymerase Chain Reactions (PCR) with the Phusion DNA Polymerase enzyme (Thermo Scientific, Wilmington, Delaware, United States of America). The PCR primers contained restriction enzyme sites were designed to introduce specific modifications according to codon usage in *E. coli* (Table 1). The PCR fragments were digested with *Bsp*HI and *Hind*III (New England Biolabs, Ipswich, MA, United States of America) and cloned into the pHAT2 vector (PERÄNEN et al., 1996), which was previously digested with *Nco*I and *Hind*III (New England Biolabs) and dephosphorylated with shrimp alkaline phosphatase (New England Biolabs). The cloned pHAT2-*orf17* and pHAT2-*est30* constructs were transformed into chemically competent *E. coli* BL21 (DE3) cells.

**Table 1 - Forward (F1) and reverse (R2) primers designed to amplify the *orf17* and *est30* genes, respectively. The sites for the restriction enzymes are underlined.**

Protein	Primers	Modifications	Enzymes sites
<b>ORF17</b>	F1 5' TATATTCATGACACAGCAGCAAGCGAAGGCG 3' R2 5' TATATAAGCTTATTCGGCGGGCGCCTTGCGTC 3'	2 <sup>nd</sup> residue: ACC – ACA (Thr)	<i>Bsp</i> HI <i>Hind</i> III
<b>Est30</b>	F1 5' ATATATCATGACAGCAAAGACCCCACTCGATCCC 3' R2 5' TATATAAGCTTACTGGCCCCGTCACGCGGCGAAG 3'	2 <sup>nd</sup> residue: TCC (Ser) – ACA (Thr) 6 <sup>th</sup> residue: CCG – CCA (Pro)	<i>Bsp</i> HI <i>Hind</i> III

### 3.3 Construction of mutants

Site-directed mutagenesis was used to introduce single point mutations (resulting in Ser to Ala mutations) in the catalytic triads of the *orf17* and *est30* genes. In addition, to analyze the relevance of the N-terminal region in Est30, two mutants were generated in which 18 (starting at Asp19) or 33 (starting at Ala34) residues were removed from the N-terminal. The primers used in this step are presented in Table 2 and the PCRs were performed using standard protocols. After digestion, the PCR products were transformed into chemically competent *E. coli* BL21 (DE3) cells. The presence of mutations was verified by DNA sequencing and the mutated enzymes were expressed and purified.

**Table 2 - Primers designed to introduce specific mutations in the studied proteins.**

	Primers	Enzymes sites	Mutation description
<b>ORF17M1</b>	F 5'GTGATGGGCATTGCCATGGGCGGCATGATC 3'	----	Ser to Ala mutation
<b>ORF17M2</b>	R 5' GATCATGCCGCCCATGGCAATGCCATCAC 3'		
<b>Est30M1</b>	F 5' GCGGCGAAGCCGCGGGCGGGCATCTC 3'	----	Ser to Ala mutation
<b>Est30M2</b>	R 5' GAGATGCCCGCCCCGCGGCTTCGCCGCC 3'		
<b>Est30asp19</b>	F 5' ATATACCATGGACGAGACGCGGGGC 3'	<i>Nco</i> I	Starting at Asp19
<b>Est30ala34</b>	F 5' ATATACCATGGCGGTGCCCGAATGG 3'	<i>Nco</i> I	Starting at Ala34

### 3.4 Expression tests

All new constructs were used for protein expression and to test the protein solubility. Therefore, one single colony of transformed cells was grown in 6 mL of YT medium containing ampicillin ( $100 \mu\text{g mL}^{-1}$ ) at  $37^\circ\text{C}$  until an  $\text{OD}_{600}$  of 0.5-0.6. Next, the culture was divided into 1 mL aliquots, which were used for the protein expression test with varying concentrations of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (0.1 mM and 0.4 mM) and temperatures ( $20^\circ\text{C}$  and  $37^\circ\text{C}$ ).

After expression for 3 hours at  $37^\circ\text{C}$  or for 20 hours at  $20^\circ\text{C}$ , the induced cultures were spun down at maximum speed and the supernatants were discarded. The cell pellets were resuspended in 100  $\mu\text{L}$  of BugBuster reagent (Novagen, Gibbstown, New Jersey, United States of America) with 0.5  $\mu\text{L}$  lysonase and incubated for 10 minutes. The lysates obtained were centrifuged and the supernatants (the soluble cell fraction) were transferred to clean tubes, whereas the pellets were resuspended with 20  $\mu\text{L}$  of BugBuster reagent. The soluble and pellet fractions were loaded onto an SDS-PAGE gel for further analysis.

### 3.5 Overexpression and purification

The transformed cells were grown in Luria Bertani (LB) or YT medium containing ampicillin ( $100 \mu\text{g mL}^{-1}$ ) at  $37^\circ\text{C}$  until the  $\text{OD}_{600}$  reached 0.5-0.6. The protein expression of *ORF17*, *Est30* and the mutants occurred with 0.4 mM IPTG at  $20^\circ\text{C}$  for 20 hours followed by centrifugation of the cultures at  $8000 \times g$  for 10 minutes at  $4^\circ\text{C}$ . The *E. coli* cells were resuspended in 30 mL of 50 mM Tris-HCl pH 8.0 and the cells were lysed with Emulsiflex (Avestin Inc, Mannheim, Germany) or with ultrasonication for 10 cycles of 10 pulses with a 40% amplitude at 10-second intervals using an LB-750 sonicator (Labometric, Leiria, Portugal). 500 mM NaCl, 10 mM imidazole, 10 mM  $\beta$ -mercaptoethanol, and 10% glycerol (v/v) was added to the material, and the crude extract was centrifuged at  $17,000 \times g$  for 45 minutes at  $4^\circ\text{C}$  to pellet the cell debris.

The supernatants containing the His-tagged recombinant proteins were loaded onto Ni-NTA resin (Qiagen, Hilden, Germany) that was previously equilibrated with purification buffer [50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 10% glycerol (v/v)]. The columns containing 3 mL of Ni-NTA resin were washed with the purification buffer containing a stepwise imidazole gradient. The *ORF17*, *Est30* and the mutants were eluted separately in purification buffer containing 200 mM imidazole.

The peak fractions containing the target protein were concentrated by centrifugation at  $3000 \times g$  at  $4\text{ }^{\circ}\text{C}$  with an Amicon Ultra-15 concentrator (Merck Millipore, Billerica, Massachusetts, United States of America). The concentrated fractions were then used for gel filtration chromatography in a HiLoad 16/60 Superdex 75 column (GE Healthcare Bio-Sciences, Uppsala, Sweden). This purification was performed in buffer containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl and 5% glycerol. The protein purity was analyzed by SDS-PAGE and the protein concentration was measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, United States of America) at the 280 nm absorption peak using the calculated extinction coefficient.

### **3.6 Circular dichroism of Est30**

To verify the influences of pH and temperature on the stability of Est30, this protein was diluted in different buffers and submitted to spectroscopic assays. The circular dichroism measurements were performed in a JASCO J-810 spectropolarimeter using 5 mM sodium phosphate pH 7.0, 5 mM sodium phosphate pH 8.0, 5 mM CHES pH 9.0, and 5 mM CHES pH 10.0 as the different buffering agents. The data were recorded from 30 accumulations obtained at  $10\text{ }^{\circ}\text{C}$  and from 195 to 260 nm, with a speed of 50 nm/min and bandwidth of 1.0 nm. For each pH experiment, a thermal unfolding was performed from  $20\text{ }^{\circ}\text{C}$  to  $105\text{ }^{\circ}\text{C}$ , allowing the temperature to equilibrate before each spectrum was recorded. The  $T_m$  represents the temperature at the midpoint of the unfolding transition. The experiments were performed using 0.2 mg/mL, which corresponded to 5.55  $\mu\text{mol}$  of the Est30 protein. All collected data were corrected for the buffer baseline contribution.

### **3.7 pH influence on the Est30 protein in the tributyrin assay**

The pure Est30 protein obtained from size-exclusion chromatography was tested at different pH values (7.0, 8.0 and 9.0) against tributyrin in Petri dishes containing 0.23% Tris (w/v), agar 1.2% (w/v) and 1% tributyrin (v/v). 20  $\mu\text{l}$  of Est30 at a concentration of 50  $\mu\text{g/ml}$  were added to each dish. The material was incubated at  $37\text{ }^{\circ}\text{C}$  for 1 day.

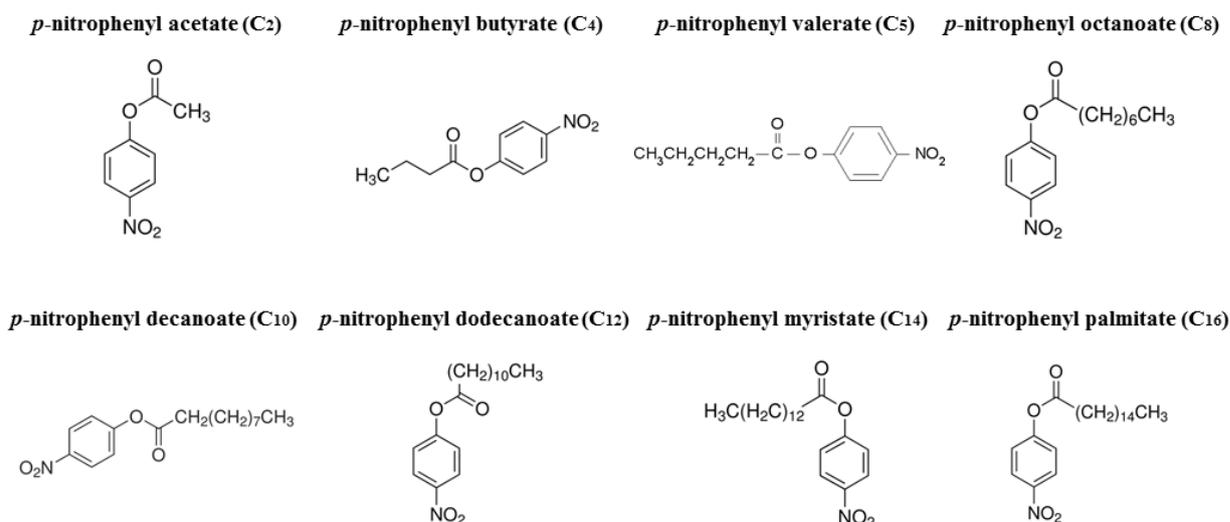
### 3.8 Functional characterization of Est30

The functional characterization of Est30 was performed by measuring the *p*-nitrophenol formation after hydrolysis of the *p*-NP acyl esters. The reactions were measured toward an enzyme-free blank to subtract the auto-hydrolysis from the results, and the released *p*-nitrophenol was measured by spectrophotometric quantification in an Enspire™ Plate Reader (PerkinElmer, Waltham, Massachusetts, United States of America) at 405 nm ( $\epsilon_{\text{pNP}} = 17.000 \text{ M}^{-1} \cdot \text{cm}^{-1}$ ). All assays were conducted in triplicate in 96-well plates with a final volume of 100  $\mu\text{L}$ . Unless otherwise indicated, the reactions were developed at 25 °C using 15.6 nM of the enzyme and 1 mM of substrates in 50 mM Tris-HCl pH 8.0 containing 0.3% triton X-100 (v/v).

To verify the enzyme activity of the purified Est30, this enzyme was tested against different *p*-NP acyl esters [acetate (C<sub>2</sub>), butyrate (C<sub>4</sub>), valerate (C<sub>5</sub>), octanoate (C<sub>8</sub>), decanoate (C<sub>10</sub>); dodecanoate (C<sub>12</sub>); myristate (C<sub>14</sub>); and palmitate (C<sub>16</sub>), Sigma-Aldrich, San Louis, Missouri, United States of America] (Figure 2). The pH optimum of Est30 was investigated using buffers of different pH values in order to establish a pH range from 5 to 11 (sodium citrate pH 5.0 and 6.0; Tris-HCl pH 7.0, 8.0 and 9.0; glycine pH 10.0; and sodium phosphate pH 11.0). 0.3% triton X-100 (v/v) was added to these reactions and *p*-NP butyrate or *p*-NP valerate were used as the substrates. The optimum temperature for Est30 activity was evaluated using *p*-NP butyrate or *p*-NP valerate and the measurements were performed by ramping the temperature 25 °C to 65 °C.

Est30 was tested for its activity in the presence of metal ions and solvents. The influence of metals (CoCl<sub>2</sub>, KCl, MgSO<sub>4</sub>, MnCl<sub>2</sub>, NiCl<sub>2</sub> and ZnCl<sub>2</sub>) was measured in the presence of 0.5 mM of each reagent and the reactions were measured against *p*-NP butyrate or *p*-NP valerate or *p*-NP octanoate. The reaction was carried out in the absence of these metal ions as a negative control. Finally, the tolerance of Est30 to solvents was assessed in the presence of 2.5 to 10% (v/v) dimethylformamide (DMF) or dimethyl sulfoxide (DMSO) against *p*-NP butyrate. The controls were solvent-free and the reactions were performed in 50 mM Tris-HCl pH 8.0.

The kinetic parameters were assessed at 25 °C using dilutions of the Est30 protein and various *p*-NP acyl ester substrates at concentrations varying from 0.0156 mM to 2 mM. The velocities of the reactions were obtained by non-linear regression of the data according to the equation  $V = V_{\text{max}} * S / (K_m + S)$ , where  $V$  is the velocity,  $V_{\text{max}}$  is the maximum enzyme velocity,  $S$  is the substrate concentration, and  $K_m$  is the Michaelis-Menten constant.



**Figure 2.  $p$ -nitrophenyl esters used for the Est30 functional characterization.**  $p$ -nitrophenyl esters from Sigma-Aldrich were used to perform the experiments, including  $p$ -nitrophenyl acetate (C<sub>2</sub>);  $p$ -nitrophenyl butyrate (C<sub>4</sub>);  $p$ -nitrophenyl valerate (C<sub>5</sub>);  $p$ -nitrophenyl octanoate (C<sub>8</sub>);  $p$ -nitrophenyl decanoate (C<sub>10</sub>);  $p$ -nitrophenyl dodecanoate (C<sub>12</sub>);  $p$ -nitrophenyl myristate (C<sub>14</sub>); and  $p$ -nitrophenyl palmitate (C<sub>16</sub>).

### 3.9 Protein crystallization and data collection

Crystallization trials of *ORF17* and Est30 were performed at the Department of Biochemistry, University of Cambridge using the following commercial kits: Classic Suite (Qiagen), JCSG Suite (Qiagen), Midas (Molecular Dimensions, Altamonte Springs Florida, United States of America), Morpheus (Molecular Dimensions), PEGsI (Qiagen), pHClear I (Qiagen) and Wizard I & II (Emerald BioSystems, Bainbridge Island, Washington, United States of America). The first hits for *ORF17* were obtained using 30% PEG 6000 and 0.1 M HEPES at pH 7.0. Two conditions allowed for Est30 crystal growth: the first was 1.6 M ammonium sulfate, 0.1 M HEPES pH 7.5, and sodium chloride 0.1 M, while the second was 1.6 M ammonium sulfate and 0.1 M bicine pH 9.0. The salt and protein concentrations were refined in order to improve the quality of the crystals.

The Est30 crystals were soaked with sodium butyrate and co-crystallized with glycerol trivalerate (Sigma-Aldrich), 1-octanoyl-rac-glycerol (Sigma-Aldrich), 1-lauroyl-rac-glycerol (Sigma-Aldrich) or stearic acid (Sigma-Aldrich). Samples of 0.28 mM of Est30 were mixed with 2 mM of each ligand, prior dissolved in 99% ethanol. The crystals were then transferred into a cryosolvent consisting of reservoir solution and 20% glycerol before being flash-cooled in a cold nitrogen stream.

The Est30 diffraction data were collected on a beamline I02 at the Diamond Light Source, England, using a Pilatus 6M detector. The diffraction data for *ORF17* and Est30 (soaked with sodium butyrate) were collected on a beamline ID14 at the European Synchrotron Radiation Facility (ESRF) using a Q315r ADSC X-ray detector.

### 3.10 Data processing and structure solving and refinement

The datasets for *ORF17* (1.4 Å), free Est30 (1.7 Å) and the Est30-butyrate complex (1.65 Å) were indexed and integrated using XDS (KABSCH, 2010). As shown in Table 3, all structures were solved using the molecular replacement method in the Phaser program provided by the CCP4 package using the closest esterase and lipase homologs found in the Protein Data Bank (PDB). For *ORF17*, an ensemble of superimposed coordinates (PDB codes 2DOD, 1Q0R, 1ZOI, 3HSS and 3OM8) was used. For Est30, the structural coordinates from PDB entries 3AIO and 2C7B were used.

The coordinates and restraints for the ligands were created with the program ProdrG2 Server (SCHÜTTELKOPF; VAN AALTEN, 2004) and used as the input for PHENIX.elbow in order to create the Crystallographic Information File (CIF) library files containing the bond lengths and angle restraints. The models were subjected to rigid body and restrained refinement cycles using the program REFMAC5, followed by several iterative rounds of refinement using refmac5 and phenix.refine (MURSHUDOV et al., 2011). COOT was used to manually correct the structures, build missing parts, and add solvent molecules (EMSLEY; COWTAN, 2004). All of the statistics are shown for model building and final refinement. The models were also subjected to the Ramachandran plot analysis for protein validation. The structural features were visualized and analyzed using Pymol (DELANO, 2010).

**Table 3 - Proteins used as templates to solve the structures of the *ORF17* and *Est30* proteins through the molecular replacement method.** The PDBs entries used to solve the *ORF17* and *Est30* structures are illustrated according to their function, their microorganism of origin, and their sequence identities (%) against *ORF17* or *Est30*.

<b>Proteins</b>	<b>Function of template structure</b>	<b>PDB code</b>	<b>Organisms</b>	<b>Identity (%)</b>
<b><i>ORF17</i></b>	Aclacinomycin methylesterase	1Q0R	<i>Streptomyces purpurascens</i>	29
	Hydrolase	3OM8	<i>Pseudomonas Aeruginosa</i> PA01	27
	Hydrolase mutant	2DOD	<i>Pseudomonas fluorescens</i>	26
	Esterase	1ZOI	<i>Pseudomonas putida</i>	28
	Protein Rv0554	3HSS	<i>Mycobacterium tuberculosis</i>	24
<b><i>Est30</i></b>	Carboxylesterase	3AIO	<i>Sulfolobus Tokodaii</i>	31
	Carboxylesterase	2C7B	Metagenomic Library	32

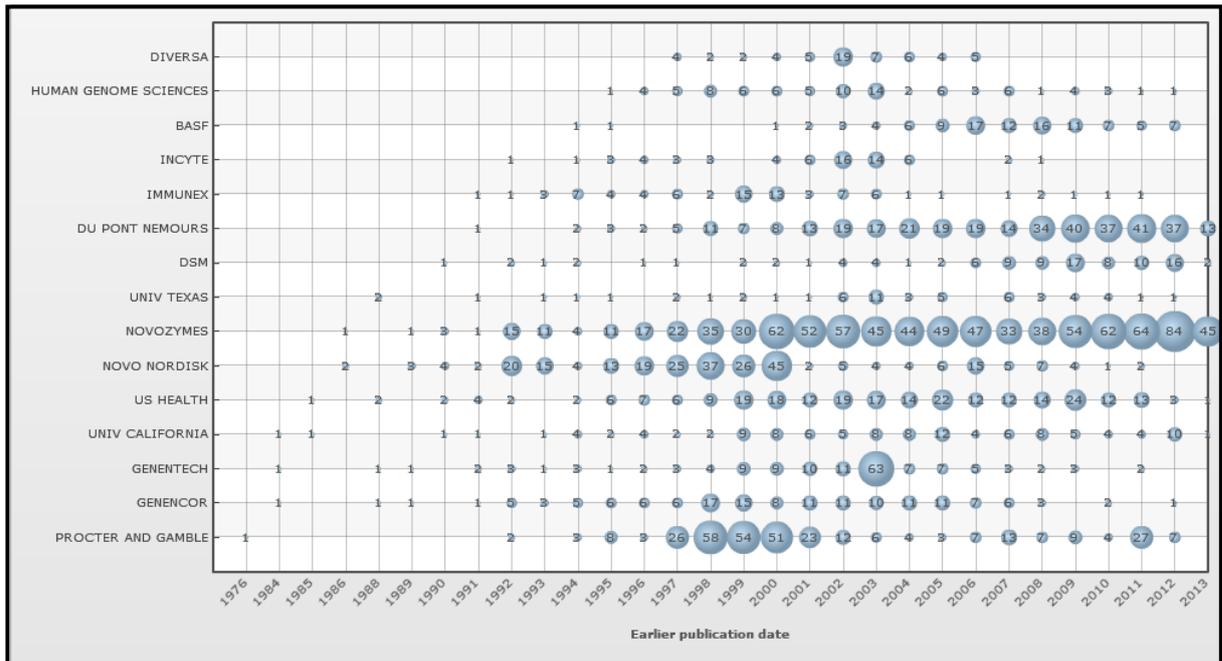
## 4 RESULTS AND DISCUSSION

### 4.1 Chapter 1: Evaluation of the global market and the search for patented sequences of esterases and lipases

Lipolytic enzymes have a high biotechnological potential due the variety of substrates upon which they can act, consequently, the range of processes in which they can be used. Once the functional characterization of four lipolytic enzymes was proposed in this project, it would be interesting to compare these enzymes to commercial and patented enzymes and to verify their biotechnological application. Thus, we performed a search for patented esterases and lipases as well as data on the global enzyme market using the Questel database. This search focused on the applicability of these enzymes, the countries and the companies of major interest.

The patent numbers of the esterases and lipases deposited according to their priority dates (first international deposit) increased significantly between the 1990s and 2000, with a peak of these patents occurring from 1996 to 2006 (Figure 3). This boost was due to the emergence of new assays for isolating and purifying these enzymes, new sequencing techniques, and mainly the development of new approaches such as the metagenomic approach. By looking at the most significant countries that have deposited esterase and lipase patents until 2013, we noted the importance that The United States of America, Europe, Australia and Canada exhibited (Figure 4) relative to the insignificance of Brazil, which holds the ninth position in number of deposited patents. Novozymes, Novo Nordisk, Procter and Gamble, and DuPont Nemours showed the largest number of patents in this field (Figure 5), though it should be noted that the continued interest of companies in these enzymes indicates that the field is not yet saturated. One example is the increase from 13 patents in 2001 to 37 patents in 2012 deposited by Du Pont Nemours.



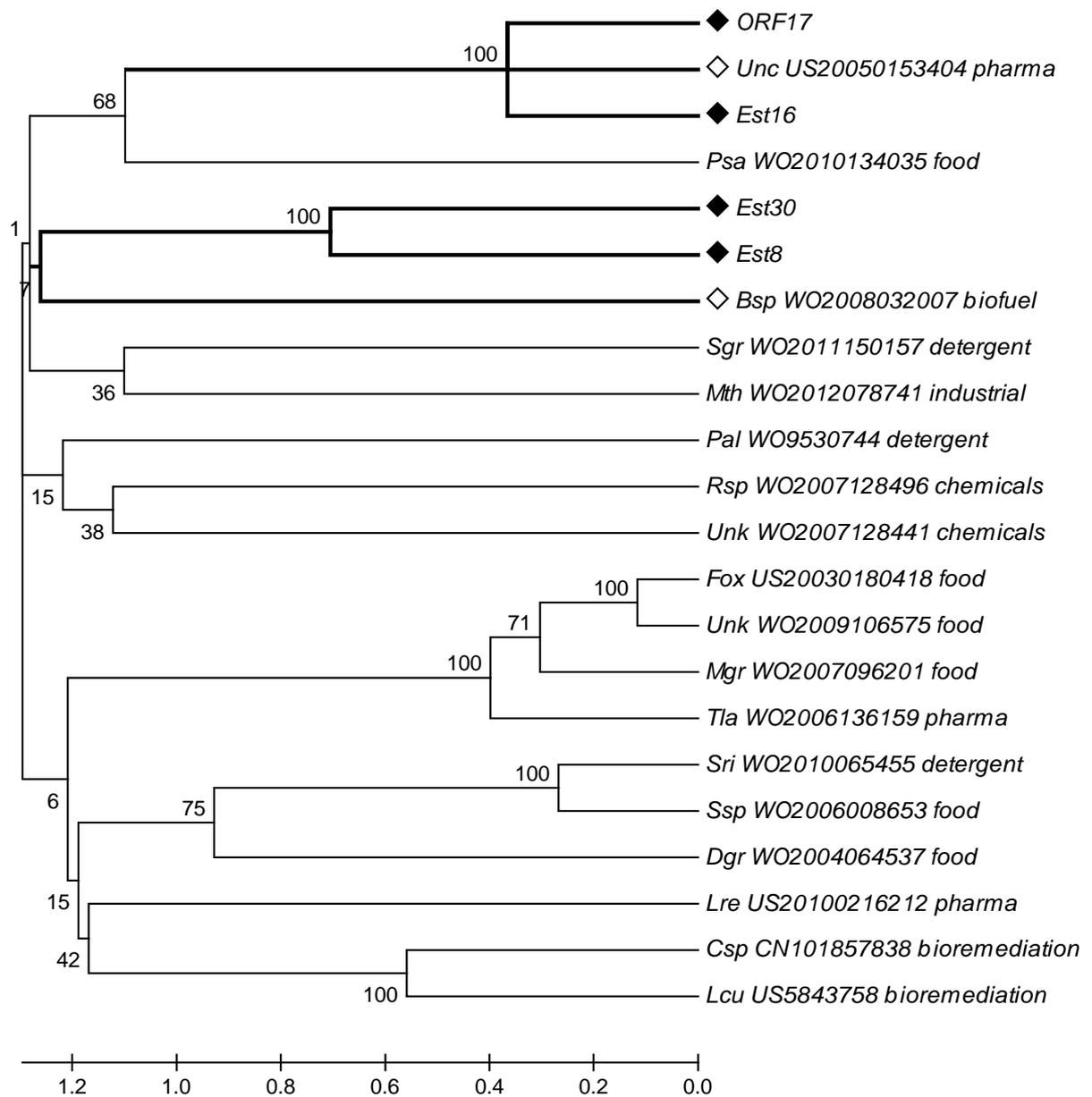


**Figure 5. Number of esterase and lipase patents deposited during the period from 1976 to 2013.** This result was obtained through an analysis conducted using the Questel Intellectual Property Portal (<http://www.orbit.com/#WelcomePage>) in March, 2013.

Through this analysis, it was possible to confirm that there has been a hot market for lipolytic enzymes and, despite the large number of deposited esterases and lipases, this number is still increasing. Advances in molecular biology techniques such as site-directed evolution, protein engineering, and the improvement of metabolic pathways were also extremely important for this increased interest, which opened opportunities for the development of new enzymes. In the context of this study, we selected 18 esterases/lipases with patented sequences extracted from GenomeQuest (Annex A), which have been used by companies and industries for different applications (including baked products, detergents, food, fine chemicals, bioremediation, waste treatment, biofuels, and others). The phylogenetic relationship between these 18 enzymes and those introduced in this study was analyzed in a phylogenetic tree after the alignment of the amino acid sequences (Figure 6).

First of all, we found that the classification was not based on the protein function and that Est16 and *ORF17* occupy a branch containing an enzyme from *Pseudomonas fluorescens* (patent number US20050153404) that has been used to prepare enantiomer-enriched esters and enantioselectivities of racemization catalysts. On the other hand, Est8 and Est30 are located in another branch containing a lipase from *Bacillus* sp. that has been applied in transesterification reactions to produce biodiesel (patent number WO2008032007) (Figure 6).

Although the number of sequences was not extensive, the proteins chosen were representatives from different sectors and applications. Thus, the results enabled the prediction of at least two types of possible applications for the enzymes as either pharmaceuticals or in biofuel production.



**Figure 6. Phylogenetic tree of Est8, Est16, ORF17, Est30, and patented esterases/lipases.** The 18 patented esterases/lipases sequences were identified using the Questel Intellectual Property Portal (<http://www.orbit.com/#WelcomePage>) and extracted from GenomeQuest (<http://www.genomequest.com/>). The sequences were aligned using the ClustalW program and the phylogenetic tree was generated using the neighbor-joining algorithm (MEGA4.0 software). The numbers at the nodes indicate the bootstrap percentages of 1000 replicates. The proteins of study (◆) and the closely related patents (◇) are shown.

## 4.2 Chapter 2: Est16 protein

In this chapter, all the results obtained for Est16 that are shown in the manuscript submitted to the Plos One Journal for analysis and publication are presented. The co-authors' authorization for the inclusion of this chapter is presented in Annex B.

### **“Est16, a New Esterase Isolated from a Metagenomic Library of a Microbe Consortium Specializing in Diesel Oil Degradation”**

Mariana Rangel Pereira<sup>1,2,3</sup>; Gustavo Fernando Mercaldi<sup>1,4</sup>; Thaís Carvalho Maester<sup>2,3</sup>; Andrea Balan<sup>1,5</sup>; Eliana G. de Macedo Lemos<sup>3\*</sup>

<sup>1</sup>National Laboratory of Biosciences (LNBio); Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo State, Brazil;

<sup>2</sup>University of São Paulo, São Paulo, São Paulo State, Brazil;

<sup>3</sup>Department of Technology, São Paulo State University, Via de Acesso Prof. Paulo Donato Castellane S/N, Jaboticabal, São Paulo State, Brazil;

<sup>4</sup>Institute of Biology, University of Campinas, Campinas, São Paulo State, Brazil;

<sup>5</sup>Department of Microbiology, Biomedical Sciences Institute II, University of São Paulo, São Paulo, São Paulo State, Brazil.

\*Corresponding author.

E-mail: egerle@fcav.unesp.br (EGML)

#### **4.2.1 Abstract**

Lipolytic enzymes have attracted attention from a global market because they show enormous biotechnological potential for applications such as detergent production, leather processing, cosmetics production, and use in perfumes and biodiesel. Due to the intense demand for biocatalysts, a metagenomic approach provides methods of identifying new enzymes. In this study, an esterase designated as Est16 was selected from 4224 clones of a fosmid metagenomic library, revealing an 87% amino acid identity with an esterase/lipase (accession number ADM63076.1) from an uncultured bacterium. Phylogenetic studies showed that the enzyme belongs to family V of bacterial lipolytic enzymes and has sequence

and structural similarities with an aryl-esterase from *Pseudomonas fluorescens* and a patented Anti-Kazlauskas lipase (patent number US20050153404). The protein was expressed and purified as a highly soluble, thermally stable enzyme that showed a preference for basic pH. Est16 exhibited activity toward a wide range of substrates and the highest catalytic efficiency against *p*-nitrophenyl butyrate and *p*-nitrophenyl valerate. Est16 also showed tolerance to the presence of organic solvents, detergents and metals. Based on molecular modeling, we showed that the large alpha-beta domain is conserved in the patented enzymes but not the substrate pocket. Here, it was demonstrated that a metagenomic approach is suitable for discovering the lipolytic enzyme diversity and that Est16 has the biotechnological potential for use in industrial processes.

#### **4.2.2 Introduction**

The global trade of industrial enzymes has been estimated to be 2.3 billion dollars, with the main profit divided into detergents (\$ 789 million), food applications (\$ 634 million) and agriculture (\$ 237 million) [1]. Currently, the enzyme market has been estimated to be € 3.4 billion with an annual growth of 6.5 to 10% [2]. Within this market, lipolytic enzymes have attracted enormous attention due to their wide biotechnological applications. They are members of the broad family of proteins containing an alpha/beta hydrolase fold and can be classified according to their substrate preferences as lipases (EC 3.1.1.3) that hydrolyse water-insoluble long-chain acylglycerols ( $C > 10$ ) or esterases (EC 3.1.1.1) that hydrolyse water-soluble short-chain acylglycerols ( $C \leq 10$ ).

Lipolytic enzymes are involved in catalyzing esters hydrolysis, ester synthesis, transesterification and other reactions. They can be used in fine chemical syntheses [3] or as flavor compounds [4,5], antioxidants and perfumes [6]. Moreover, esterases can show high regio- and stereospecificity, can hydrolyse a wide range of substrates, do not require the presence of cofactors, and can be stable and active in organic solvents [7,8].

In view of the high demand for biocatalysts in biotechnology, new experimental approaches have been developed over the last few years to search for efficient molecules. Metagenomics is a cultivation-independent method that gives access to the collective genome [9] and to uncultured microorganisms, which is a vast source for gene exploration in biotechnology. In recent years, several genes encoding lipolytic enzymes were identified in metagenomic libraries from different environmental samples [10-16]. Additionally, lipolytic enzymes that display resistance to organic solvents [8,17] or salt tolerance [18] and show high

potential for use in industrial processes could be prospected through metagenomic approaches.

Despite the interesting discoveries of new lipolytic enzymes from the metagenome, most enzymes remain uncharacterized. From the first screening for lipolytic enzymes in 2000 until 2009, 76 positive clones of new esterases or lipases were identified, but only 11 were overexpressed and purified for biochemical characterization [19]. Here, the cloning, overexpression, functional characterization and preliminary structural analysis of a novel esterase, Est16, is reported. Est16 was isolated from a metagenomic library of a microbe consortium specializing in diesel oil degradation. The results showed that Est16 is a new esterase member of Family V of bacterial lipolytic enzymes and shares some structural similarities with an aryl-esterase from *Pseudomonas fluorescens* [20] and a patented Anti-Kazlauskas lipase (patent number US20050153404). Est16 was expressed in *Escherichia coli* cells as a soluble, monodisperse and highly stable protein. Kinetics assays revealed that Est16 is active against a broad range of substrates with short and long acyl chains, but has the highest catalytic efficiency against *p*-nitrophenyl butyrate (C<sub>4</sub>) and *p*-nitrophenyl valerate (C<sub>5</sub>). In addition, the enzyme activity was maintained at different pHs and in the presence of detergents, organic solvents and ions. Altogether, these data show that Est16 presents a set of characteristics that make it an interesting candidate for biotechnological applications.

### **4.2.3 Methods**

#### **4.2.3.1 Sampling**

In a previous work, a soil sample was collected from a region contaminated with petroleum hydrocarbons from a factory of automotive lubricants in the city of Ribeirão Preto, São Paulo State, Brazil [21]. The soil was used to develop a microbe consortium at the Department of Biology at the University of São Paulo State, Jaboticabal Campus [21], which had the bacterial diversity analyzed [22]. From this microbe consortium, a metagenomic library of fosmid vectors was built in the Biochemistry Laboratory of Microorganisms and Plants (LBMP) in the Department of Technology at the same university. In this work, the 4224 clones obtained were used as source for prospecting lipolytic enzymes.

#### 4.2.3.2 Screening for lipolytic activity

The fosmid clones from the DNA metagenomic library were submitted to lipolytic activity screening on Luria Bertani (LB) agar Petri dishes containing 1% tributyrin, 1% gumarabic and  $12.5 \mu\text{g mL}^{-1}$  chloramphenicol. Arabinose (0.001%) was added to induce a high copy number of the plasmids based on the oriV/trfA system [23]. As a negative control, *E. coli* EPI300™ cells (Epicentre, Technologies, Madison, WI, USA) carrying the pCC2FOS vector supplied by a kit were used. The plates were incubated at 37 °C for 3 days and then were transferred to 4 °C for up to 7 days. The presence of a halo around the colonies was indicative of substrate hydrolysis and lipolytic activity by the clones, which were then replicated to confirm the result.

#### 4.2.3.3 Selection of the clone and sub-library construction

A fosmid clone that formed a clear halo was selected for construction of the sub-library and to identify the proper gene encoding the putative esterase/lipase. The DNA clone was fragmented using the nebulization method [24] and the material was applied to a 1% agarose gel for electrophoresis. The 1 to 3 kb fragments of interest were recovered, purified from the gel with the GeneClean III kit (BIO 101 Inc., Vista, CA, USA), and cloned into the pUC19 vector (Fermentas, Burlington, ON, Canada), which was previously digested with *Sma*I (Fermentas) and dephosphorylated (New England Biolabs, Ipswich, MA, USA). The ligation mixture was used for transformation of *E. coli* DH5 $\alpha$  cells and the selection of the clones was performed as previously described using 96 wells plates supplemented with LB media and  $70 \mu\text{g mL}^{-1}$  ampicillin [24].

#### 4.2.3.4 DNA sequencing, gene annotation and qualitative analysis of the *Est16* activity

The plasmid DNA extracted from the subclones of the sub-library was sequenced in both directions using the M13-Forward and M13-Reverse primers. For sequencing the fosmid ends, the pCC2™F and pCC2™R primers (Epicentre) were used. Sequencing was performed on an ABI 3100 instrument (Applied Biosystems, Foster City, CA, USA) and the electropherograms were analyzed with the aid of the PhredPhrap program [25, 26]. The contig was compared with the sequences deposited in the National Center for Biotechnology Information (NCBI) using the local alignment BLAST tool [27] against the nucleotide

collection (nr/nt). Gene annotation was performed using the ORF Finder from NCBI (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) and the prediction function was identified using a BLASTX query against the non-redundant bank protein (nr), taking into account the maximum score, the query coverage and the E-value. To check if the subclone containing the putative esterase/lipase gene presented lipolytic activity, bacterial colonies were plated onto Petri dishes containing LB-agar with 1% tributyrin, 1% gumarabic and  $70 \mu\text{g mL}^{-1}$  ampicillin, and incubated at  $37^\circ\text{C}$  for 3 days and then  $4^\circ\text{C}$  for 4 days. As mentioned before, the activity was evidenced by the appearance of a clear halo around the colonies on the plates.

#### 4.2.3.5 Phylogenetic analysis

Two distinct phylogenetic trees were built to compare the phylogenetic relationship of Est16 with the 34 previously described esterases/lipases from the eight families proposed by Arpigny and Jaeger [28] and with 18 sequences from patented enzymes that show industrial applications. Sequences of the 34 esterases/lipases were obtained from NCBI and were then aligned with Est16 using the ClustalW program [29] (Table S1). The alignment file was submitted to the Mega 4.1 program [30] for building the phylogenetic tree based on the neighbor-joining algorithm [31].

The choice of patented sequences was based on the literature data [32] and industrial applications using the Orbit database at the Questel website (<https://www.questel.com>). After the search, the sequences were obtained from the GenomeQuest website (<http://www.genomequest.com/>) (Table S2) and aligned for tree preparation, which was built as mentioned before.

#### 4.2.3.6 Molecular modeling of Est16 and structural analysis

The molecular model of Est16 was built with the Modeller 9v4 program [33] using the structural coordinates of an aryl-esterase from *Pseudomonas fluorescens* (PDB code 1VA4) [20] as a model. In addition, to evaluate the structural features of Est16 in comparison with the patented Anti-Kazlauskas lipase (patent number US20050153404), a model of this structure was built using the structural coordinates of an esterase from *Pseudomonas fluorescens* (PDB code 3IA2) [34] that shares 22.8% sequence identity with the patented enzyme. Then, the superposition of the structures was performed using COOT [35]. The

sequence comparison and the structural analysis were performed using ClustalW [29] and the Pymol program [36], respectively. All final figures were prepared with Pymol.

#### 4.2.3.7 Cloning, expression and purification of *Est16*

The *est16* gene was amplified by Polymerase Chain Reaction (PCR) with the Pfu DNA Polymerase enzyme (Fermentas) using the following primers: Forward 5' CTGAATTCTCCATGCCGCAGGTTTCAG 3' and Reverse 5' GGGCTCGAGCGTCACTCCGCC 3', which had sites for the restriction enzymes *EcoRI* and *XhoI*, respectively (underlined). The amplified fragment consisting of 917 pb was previously digested with the mentioned enzymes and cloned into the pET28a vector (Novagen, Gibbstown, NJ, USA) for the protein expression. The recombinant pET28a-*est16* plasmid was used for transformation of chemically competent *E. coli* BL21 (DE3) cells. For expression tests, a pre-inoculum of the cells was grown at 37 °C for 20 hours in 5 mL of LB media containing 50 µg/mL kanamycin. This pre-inoculum was then used to inoculate 500 ml of LB (50 µg/mL kanamycin). The culture was grown under shaking at 250 r.p.m. at 37 °C until the OD<sub>600nm</sub> reached 0.5-0.6, at which point 0.1 mM isopropyl-β-D thiogalactopyranoside (IPTG) was added to induce protein expression at 28 °C for 20 hours. After this period, the cells were centrifuged at 8000 x g for 10 minutes at 4 °C and resuspended in 35 mL of lysis buffer [10 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, 10% glycerol (v/v) and 0.25% (w/v) nonidet P-40]. The material was maintained in an ice bath for 1 hour with 1 mM DL-Dithiothreitol (DTT) and 4 µg/mL of lysozyme. Ultrasonication was performed for 10 cycles of 10 pulses at a 40% amplitude with 10 second intervals using an LB-750 sonicator (Labometric, Leiria, Portugal). The obtained extract was centrifuged at 17,000 x g for 45 minutes at 4 °C to pellet the cell debris. The expression and solubility of the protein was confirmed after loading the non-induced, induced, extract and pellet samples into an SDS-PAGE gel. After the solubility of *Est16* was verified, 3 mL of Ni-NTA resin (Qiagen, Hilden, Germany), which was previously equilibrated with the purification buffer [50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole and 10% glycerol (v/v)], was added to the soluble extracts for protein purification using immobilized metal affinity chromatography (IMAC). The sample was incubated for 1 hour at 4 °C under agitation before it was loaded into a gravity flow column. After the wash of the resin with 10 column volumes of purification buffer containing 10 mM of imidazole, the protein was eluted in a stepwise imidazole gradient (50-500 mM imidazole). The eluted fractions were concentrated with

Amicon Ultra-15 filters (Merck Millipore, Billerica, MA, USA) under centrifugation at 3000  $\times g$  at 4 °C. To eliminate the contaminants, additional size-exclusion chromatography (SEC) was performed using the Hiload 16/60 Superdex 200 column (GE Healthcare Bio-Sciences, Uppsala, Sweden). This purification was performed with the buffer (20 mM Tris-HCl pH 8.0; 50 mM NaCl; 5% glycerol) at 20 °C. All samples were analyzed using SDS-PAGE and the protein concentration was measured using a Nanodrop ND-1000 spectro-photometer (Nanodrop Technologies, Wilmington, DE, USA).

#### 4.2.3.8 Tributyrin assay in Petri dishes

The pure protein obtained from size-exclusion chromatography was tested in different pHs (7.0, 8.0 and 9.0) against tributyrin in Petri dishes containing 0.23% Tris (w/v); agar 1.2% (w/v); and 1% tributyrin (v/v). Est16 was prepared at a concentration of 50  $\mu\text{g/ml}$  and 20  $\mu\text{l}$  were added to each dish, after, they were incubated at 37 °C for 1 day.

#### 4.2.3.9 Measuring the enzymatic characteristics of Est16

The Est16 activity was determined via hydrolysis of *p*-nitrophenol (*p*-NP) ester substrates by measuring the formation of *p*-nitrophenol. The amount of *p*-nitrophenol released was quantified by absorbance at 405 nm ( $\epsilon_{\text{pNP}} = 17.000 \text{ M}^{-1} \cdot \text{cm}^{-1}$ ) using a 2104 Envision Multilabel Reader (PerkinElmer, Waltham, Massachusetts, USA). All assays were carried out in triplicate in 96-well plates with a final volume of 100  $\mu\text{L}$ . Pipetting steps were performed using a Janus Varispan automated liquid handler (PerkinElmer). Unless otherwise indicated, the reactions were carried out at 30 °C using 15 nM of the enzyme and 1 mM of substrate in 50 mM Tris-HCl pH 8.0 and 0.3% triton X-100 (v/v). For the analysis of the substrate specificity range of the purified Est16, the hydrolysis of the following *p*-NP-fatty acyl esters (Sigma, MI, USA) was measured: butyrate ( $\text{C}_4$ ), valerate ( $\text{C}_5$ ), octanoate ( $\text{C}_8$ ), decanoate ( $\text{C}_{10}$ ), dodecanoate ( $\text{C}_{12}$ ), myristate ( $\text{C}_{14}$ ) and palmitate ( $\text{C}_{16}$ ). The optimum temperature for the enzyme activity was evaluated using *p*-NP valerate and *p*-NP octanoate by measuring the production of *p*-nitrophenol in the range of 25 °C to 65 °C (5 °C steps) for 10 minutes using an Enspire™ Plate Reader (PerkinElmer). The effect of the pH on the activity of Est16 against *p*-NP butyrate was measured at pH 5.0 to 11.0 using the following buffers at 50 mM: sodium citrate pH 5.0 and 6.0; Tris-HCl pH 7.0, 8.0 and 9.0; N-cyclohexyl-2-

aminoethanesulfonic (CHES) pH 10.0; and sodium phosphate pH 11.0. All reactions contained 0.3% triton X-100 (v/v).

The influence of metals ions on the performance of Est16 was evaluated using the substrate *p*-NP butyrate, and the activity was measured in the presence of 0.5 mM of the following reagents: CaCl<sub>2</sub>, CoCl<sub>2</sub>, KCl, MgSO<sub>4</sub>, MnCl<sub>2</sub>, NiSO<sub>4</sub>, and ZnCl<sub>2</sub>. An assay with 0.5 mM of ethylenediaminetetraacetic acid (EDTA) was also performed. Controls of the reactions were obtained for all of the reagents except the metals, ions and EDTA. Lastly, to evaluate the tolerance of the enzyme to detergents and solvents, the Est16 hydrolysis of *p*-NP butyrate was evaluated in the presence of the detergents tween 20 and triton X-100 [0.3125 to 5% (v/v)] and with the solvents dimethylformamide (DMF) and dimethyl sulfoxide (DMSO) [2.5 to 10% (v/v)]. Neither the solvents nor the detergents were added to the control samples. The reactions were carried out in 50 mM Tris-HCl pH 8.0.

The kinetic parameters of Est16 were calculated from different reactions. The initial velocity of the reaction was measured using 15 nM of the enzyme for substrates C<sub>4</sub> – C<sub>10</sub> and 50 nM for substrates C<sub>12</sub> – C<sub>16</sub>. Two set of substrate concentrations were used depending on the chain length: 0.0078 mM to 1 mM for *p*-NP acetate, butyrate, valerate, and octanoate, and 0.0156 mM to 2 mM for *p*-NP decanoate, dodecanoate, myristate and palmitate. The kinetic parameters for each substrate were obtained by non-linear regression of the data using the following equation:  $V = V_{max} * S / K_m + S$ , where  $V$  is the velocity,  $V_{max}$  is the maximum enzyme velocity,  $S$  is the substrate concentration, and  $K_m$  is the Michaelis-Menten constant.

#### 4.2.3.10 Circular dichroism

Circular dichroism (CD) measurements were carried out with a JASCO J-810 spectropolarimeter equipped with a Peltier-type temperature controller and a thermostatic cell holder, which was interfaced with a thermostatic bath. To verify the influence of pH, the Est16 protein was diluted in different buffers (5 mM sodium phosphate pH 7.0; 5 mM sodium phosphate pH 8.0; 5 mM CHES pH 9.0) and submitted to the CD analysis. Thirty consecutive scans from 195 to 260 nm were compiled and the average spectra were stored. The data were corrected for the baseline contribution of the buffer. Thermal unfolding experiments were performed in 1-mm-long cells by increasing the temperature from 20 °C to 110 °C. The temperature was allowed to equilibrate before each spectrum was recorded. The  $T_m$  represents the temperature at the midpoint of the unfolding transition. CD of Est16 was performed at the concentration of 5.64 μmol of protein.

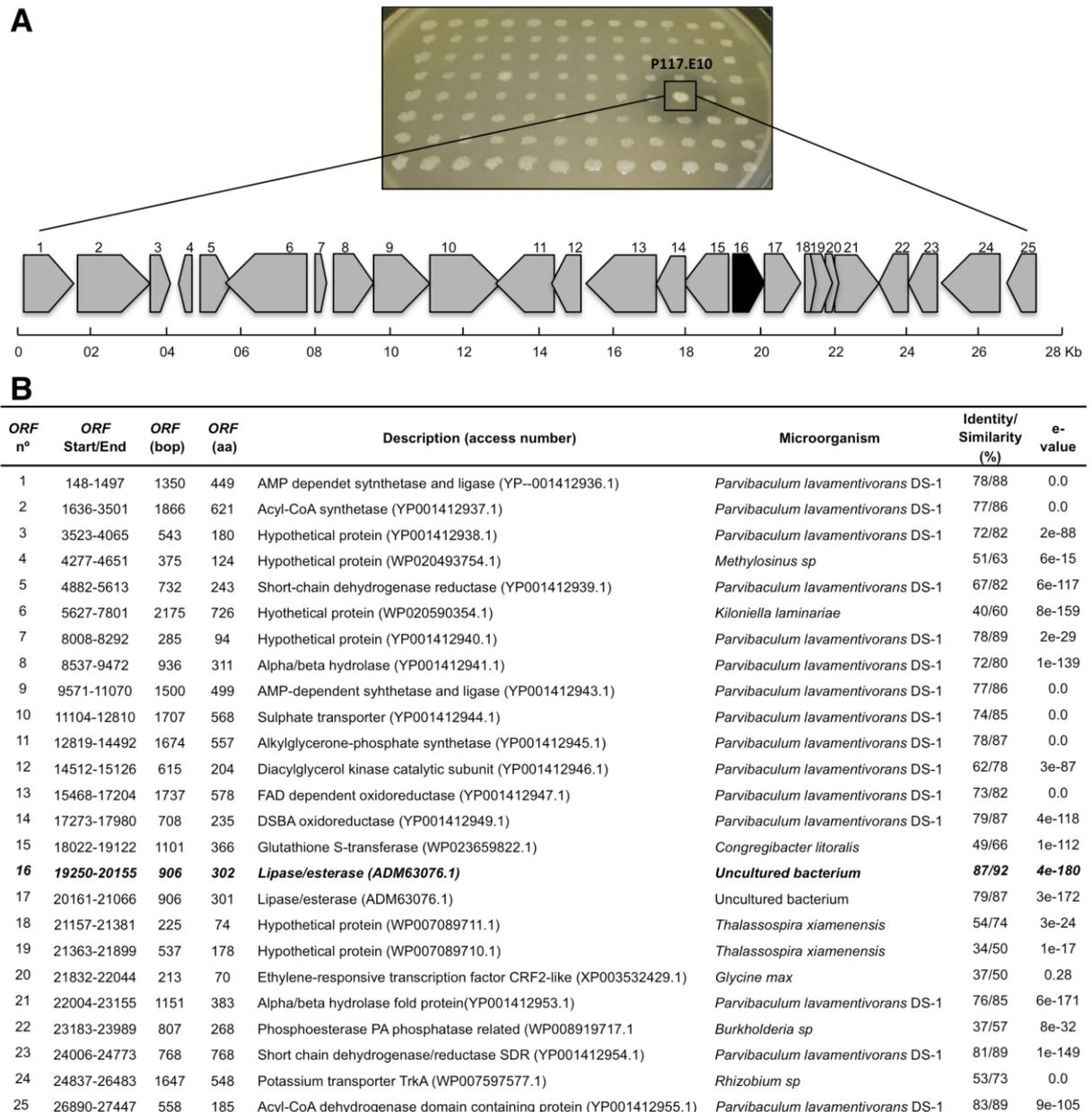
#### 4.2.3.11 Nucleotide sequence accession number

The DNA sequence of the *est16* gene was deposited in GenBank under the accession number KM882609.

### 4.2.4 Results and Discussion

#### 4.2.4.1 Metagenomic library construction and screening for lipolytic activity

A fosmid metagenomic library was constructed using high molecular weight DNA isolated from a microbe consortium specialized for the degradation of diesel oil. Approximately 4224 clones were obtained with an average insert DNA size of 30 to 40 kb, representing a total size of 148 Mb of the total metagenomic DNA. All clones were screened for lipolytic activity based on the hydrolytic activity of emulsified tributyrin (1%), resulting in 30 positive lipolytic clones that showed a clear zone around the colonies, which corresponded to approximately one lipolytic gene per 4.9 Mb DNA (one positive clone per 140 tested clones). From one of the screening plates with 96 colonies, the P117.E10 clone exhibited strong lipolytic activity as evidenced by the clear halo (Figure 1A). The P117.E10 clone was chosen for construction of the sub-library and gene annotation. The fosmidial DNA of the P117.E10 clone was extracted and an insert of approximately 35 kb was subcloned. 672 subclones with 1 to 3 kb inserts were collected and the plasmid DNA was extracted and quantified for sequencing. The final contig was 27,485 bp and shared a 78% nucleotide sequence identity and 52% query coverage with *Parvibaculum lavamentivorans* (accession number CP000774.1). From 147 ORFs identified by the ORF Finder, 25 were selected using BLASTP (Figure 1B). From these, ORF16, which consists of 906 bp (302 amino acids, in bold), was identified as a putative gene encoding an esterase/lipase that was similar to an esterase/lipase from an uncultured bacterium (accession number ADM63076.1; 87% sequence identity) and a protein from *Parvibaculum lavamentivorans* DS-1 (accession number WP012110584.1; 82% sequence identity) with alpha/beta hydrolase fold. The ORF16 was chosen for further analysis and the corresponding protein was designated Est16.



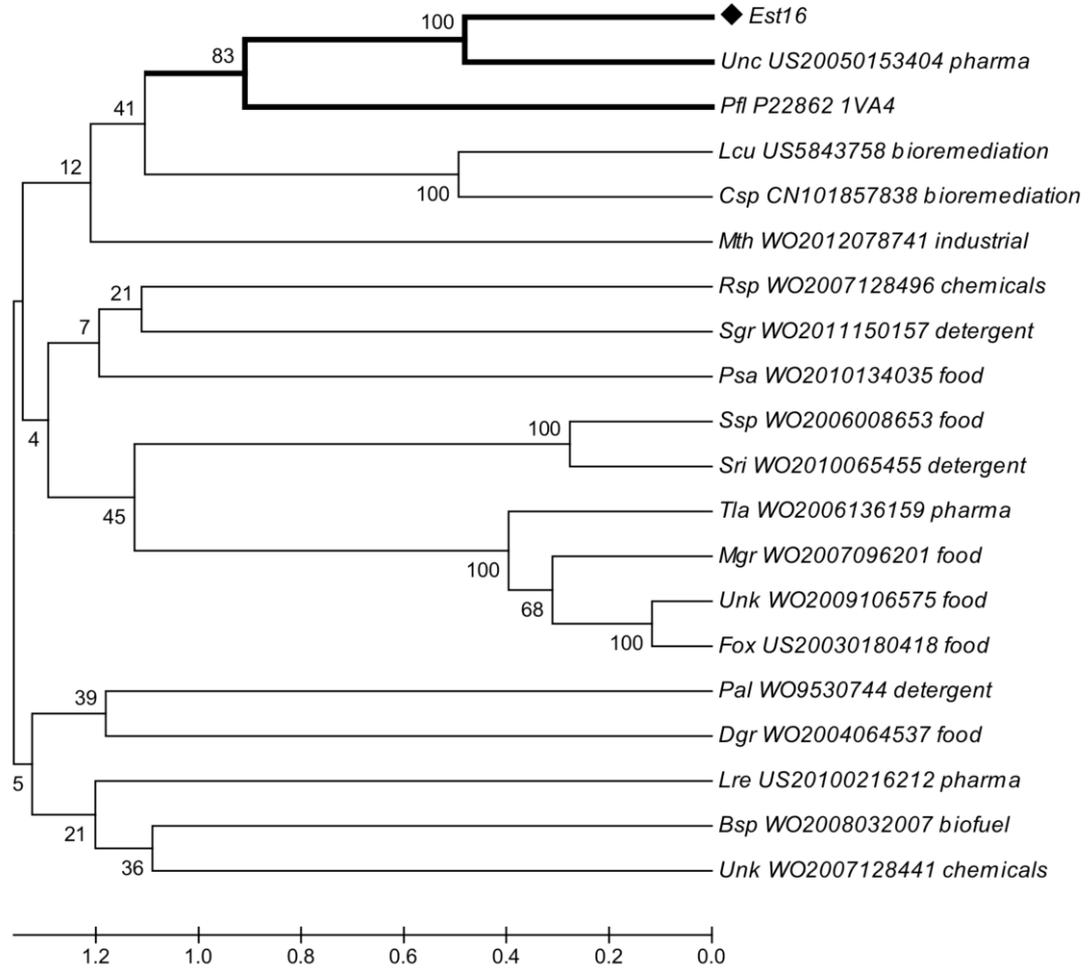
**Figure 1. Analysis of the lipolytic activity and characterization of the P117.E10 clone.** (A) Clones obtained from the metagenomic approach were cultivated in a Petri dish containing LB media supplemented with 1% tributyrin (v/v), 1% gumarabic (w/v), 0.00125% chloramphenicol (v/v) and 0.001% arabinose (v/v). The plate was maintained at 37°C for three days and for four more days at 4 °C. A clear halo was observed around the P117.E10 fosmid clone, evidencing its ability to degrade tributyrin. A detailed physical map of P117.E10 is shown, which consists of 25 identified ORFs depicted as arrows according to their location and direction in the fosmid vector. The black arrow indicates ORF16, which encodes the Est16 protein. (B) The characteristics of each ORF were obtained after using ORF Finder and the BLASTP search. The putative function of each ORF and their GenBank accession numbers are also shown.

#### 4.2.4.2 *Est16 shares sequence and structure similarities with members of family V and with the patented Anti-Kazlauskas lipase*

To first evaluate the phylogenetic relationships between Est16 and the 34 sequences from the esterases/lipase according to Arpigny and Jaeger [28], a phylogenetic tree was built using the Mega 4.1 program after the alignment of the Est16 amino acid sequence with all sequences (Figure S1). Est16 is related to a lipolytic enzyme from *Sulfolobus acidocaldarius* (accession number AAC67392.1) and an esterase/lipase from *Haemophilus influenzae* (accession number AAC21862.1), which share 29% and 22% amino acid sequence identity with Est16, respectively (Figure S1A). According to Arpigny and Jaeger [28], these enzymes belong to family V of bacterial lipolytic enzymes. Est16 multiple sequence alignment with *Sulfolobus acidocaldarius* (accession number AAC67392.1) confirmed the conservation of the consensus region of family V and the residues around the catalytic triad formed by Ser118, Asp245 and His273. The catalytic nucleophile Ser is located in the central portion of the conserved motif G-X-S-M-G-G (Figure S1B). In addition, two other enzymes that belong to family V, the triacylglycerol lipases from *Psychrobacter immobilis* (accession number CAA47949.1) and *Moraxella* sp. (accession number CAA37863.1), showed a close relationship with members of family I.

To compare Est16 with proteins with solved three-dimensional structures, we submitted the amino acid sequence of Est16 to BLASTP against the Protein Data Bank [PDB (<http://www.rcsb.org/pdb/home/home.do>)]. The results revealed that Est16 has structure similarity to an aryl-esterase from *Pseudomonas fluorescens* (PDB code 1VA4) [20] that shares a 22.7% amino acid sequence identity with Est16. To verify whether Est16 presented similarities with patented enzymes, we also built a second phylogenetic tree that included 18 sequences of esterases/lipases extracted from GenomeQuest and the sequence of the *P. fluorescens* aryl-esterase [20]. Interestingly, this second analysis corroborated the results of the BLASTP query against the PDB, revealing that Est16 is most closely related to two enzymes: these are an enzyme from *P. fluorescens*, the Anti-Kazlauskas lipase (patent number US20050153404) that is used to prepare enantiomer-enriched esters and enantioselectivities in a racemization catalyst, and an esterase from *Lucilia cuprina* (patent number US5843758) that is used as bioremediation agent (Figure 2). The tree revealed that in general, the enzymes were grouped according to their applications.

A



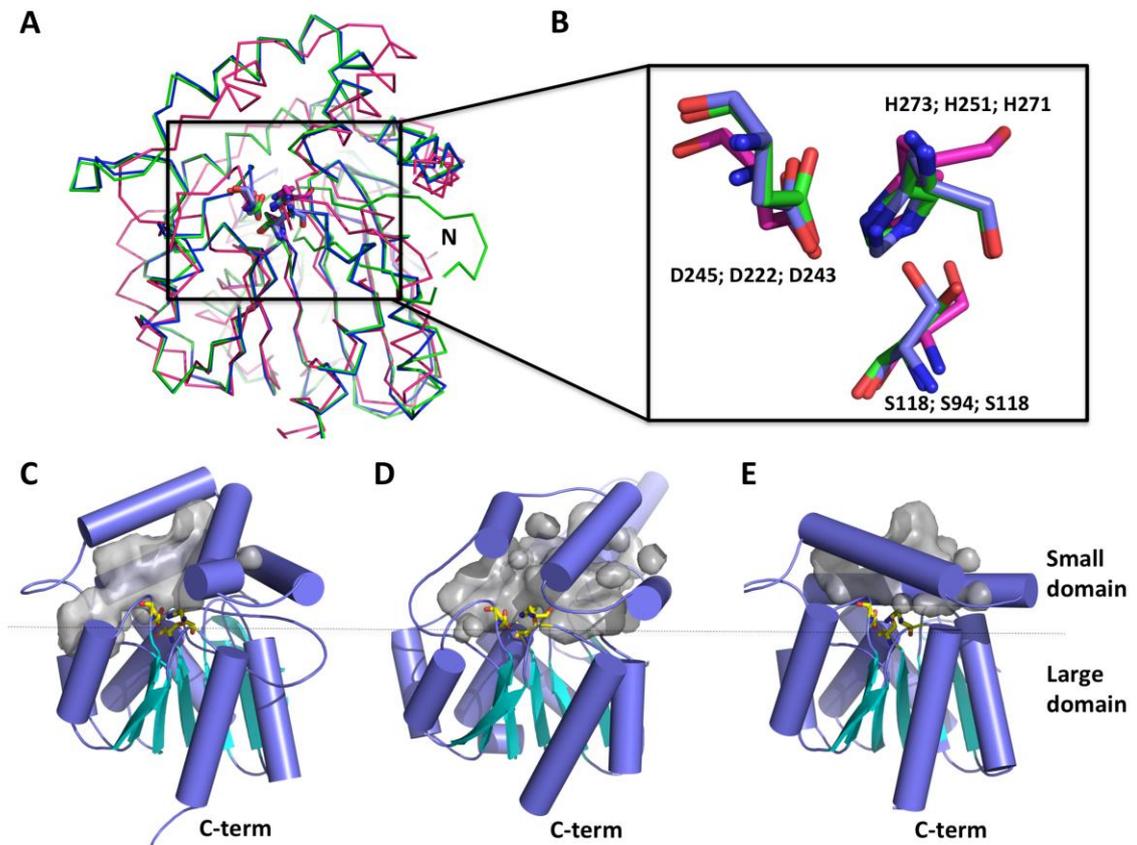
B

Consensus	GX <sup>S</sup> XGG							
<b>Est16</b>	112	AHIVG <b>A</b> SMGG	121	236	<b>PAV</b> VLHGGDD	245 264	ELIVVPGMGH	273
<b>Unc US20050153404 pharma</b>	112	THVVG <b>V</b> SMGG	121	234	<b>PTL</b> VISGAED	243 262	RFELIEGMGH	271
<b>Pfl P22862 PDB1VA4</b>	88	VTLVG <b>F</b> SMGG	97	213	<b>PTL</b> VIHGDD	222 242	ELKVYKDAPH	251
		*			*		*	

**Figure 2. Phylogenetic tree and comparison of Est16 with esterases/lipases showing their biotechnological applications.** (A) A phylogenetic tree of 18 sequences from patented esterases/lipases, the sequence of aryl-esterase (PDB code 1VA4) and Est16 was generated using the neighbor-joining algorithm (MEGA4.0 software). The numbers at the nodes indicate the bootstrap percentages from 1,000 replicates. (B) The amino acid sequence alignment of Est16 with Anti-Kazlauskas lipase (patent number US20050153404) and aryl-esterase (PDB code 1VA4) localized all three enzymes on the same same branch. The residues involved in the substrate-pocket and family classification are shown in bold and the catalytic triad residues are denoted with asterisks (\*). The consensus motif of family V is shown.

#### 4.2.4.3 The structural features of Est16 model corroborates the sequence similarities with *Pseudomonas fluorescens* and the Anti-Kazlauskas lipase

As previously shown, the BLASTP query of the Est16 amino acid sequence versus the PDB confirmed the phylogenetic analysis and revealed the best score for the three-dimensional structure of the *Pseudomonas fluorescens* aryl-esterase (PDB code 1VA4) [20]. Using the structural coordinates of this enzyme, we built a three-dimensional model of Est16. Moreover, based on the close relationship between Est16 and the Anti-Kazlauskas lipase, we also modeled the structure of this patented enzyme for further structural analyses. Est16 model was nicely superposed onto the structures of the *P. fluorescens* and Anti-Kazlauskas lipases structures, revealing r.m.s.d. values of 2.8 Å and 3.6 Å, respectively. The three structures conserved the large alpha/beta domain that characterizes a hydrolase fold and sustains the smaller alpha-helical domain, which showed significant differences between the three structures (Figure 3A). 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 in the three structures irrespective of the structural differences (Figure 3B). For comparison of the full structures, we positioned the C-termini facing down maintaining the  $\beta$ -sheets in the horizontal plane, and depicted the catalytic triad as a yellow stick in between both domains. In this way, the differences in the substrate-binding pocket (in gray surface) are evidenced (Figure 3C-E) due to the number and positioning of the  $\alpha$ -helices. The arrangement and folding of this helices altered the size and dimensions of the substrate-binding pockets (Figure 3C-E, gray surface). Compared to the Est16 substrate-binding pocket, the helices of the Anti-Kazlauskas lipase, which were almost parallel in orientation to those of Est16, formed a bigger channel compared to the small channel presented by the *P. fluorescens* enzyme. Due to these characteristics, this region can be considered important for substrate selectivity, enzymatic activity, or protein engineering using rational design or directed evolution targeting the enantioselectivity of the enzyme.

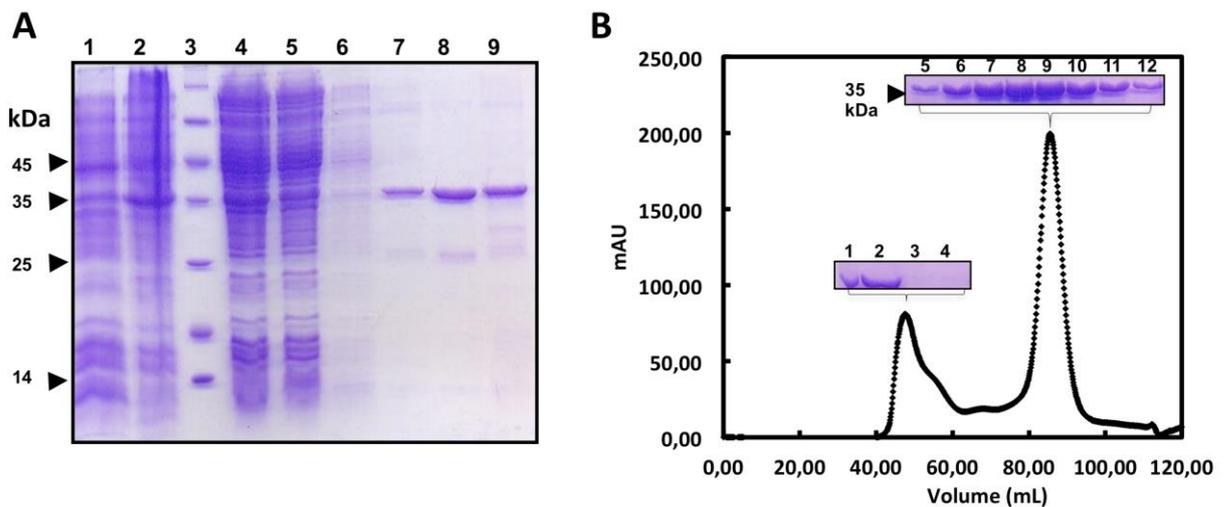


**Figure 3. Structural features of the Est16 model in comparison with an esterase from *P. fluorescens* and the Anti-Kazlauskas lipase.** (A) Ribbon representation of the Est16 model (green) superposed on the structures of *P. fluorescens* esterase (blue) and the patented lipase (patent number US20050153404) (purple) revealing the conservation of the alpha/beta fold. N is the N-terminal. (B) A detailed picture of the residues from the catalytic triad of the three enzymes showing the structural superposition. Residues are shown in the following order: Est16, *Pfl* (PDB code 1VA4) and Anti-Kazlauskas lipase. A comparative analysis of the domains and pockets of Est16 (C), the Anti-Kazlauskas lipase model (D) and the structure of the *P. fluorescens* aryl-esterase (E) are shown in the cartoon. Helices are blue,  $\beta$ -sheets are cyan and the residues in the active site are shown as yellow sticks. The large and small domains are delimited in the three proteins.

#### 4.2.4.4 Est16 was expressed and purified as a soluble, monodisperse and stable protein

The recombinant Est16 protein was expressed in the BL21 (DE3) strain of *E. coli* as a soluble cytosolic protein genetically fused at the N-terminal end to a His<sub>6</sub>-Tag and the thrombin cleavage site encoded by the pET28a expression vector. The molecular mass of the recombinant Est16 protein was 35.4 kDa, as expected from its amino acid sequence. The maximum soluble protein yields, representing up to 60% of the total expressed protein, were achieved after 20 hours of cultivation in LB media at 28 °C with aeration. The recombinant protein was first purified by immobilized metal affinity chromatography and eluted using a buffer containing 50 mM imidazole (Figure 4A), followed by one step of size exclusion

chromatography (Figure 4B). Two peaks were obtained in the second purification step, but the DLS analysis showed that the protein was monodisperse only in the second, most abundant peak. The recombinant Est16 recovered at the end of the purification was concentrated and remained soluble and stable at high concentrations (more than 10 mg/mL) even after prolonged storage at 4 °C. Its stability allowed for spectroscopic assays and further experiments to determine its kinetic parameters.

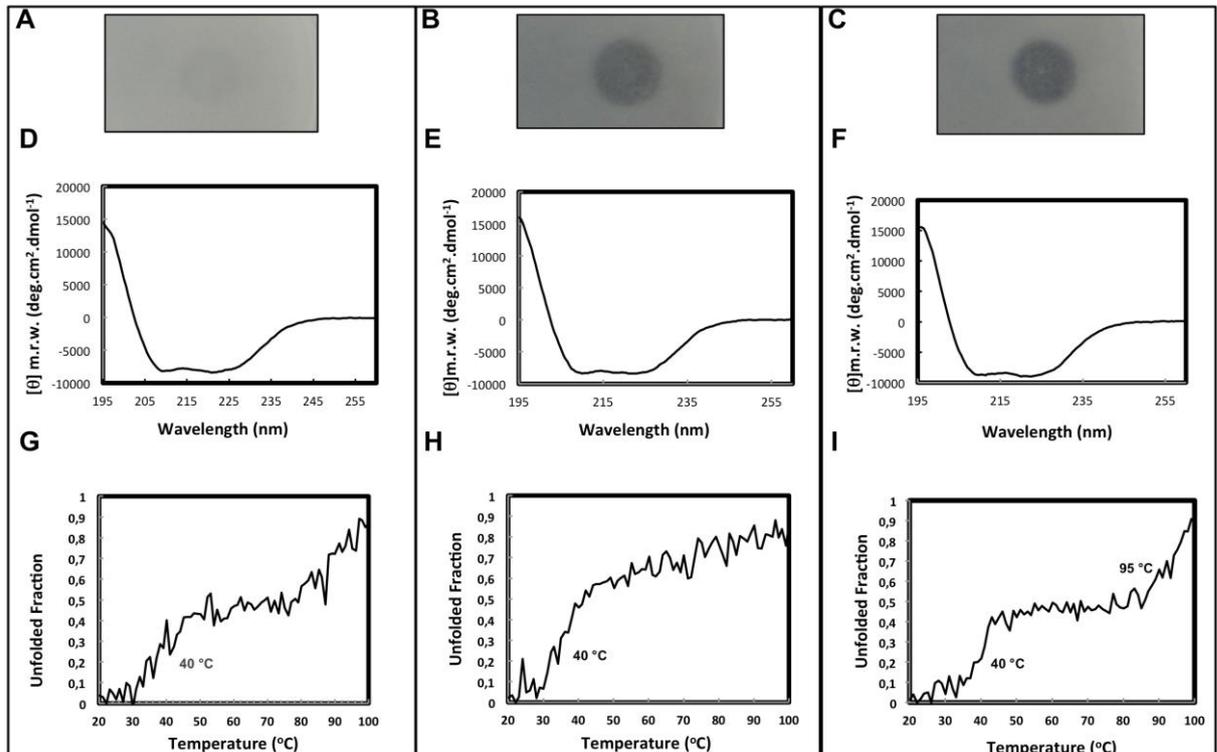


**Figure 4. Production steps of Est16 as a soluble and stable protein.** (A) 12% SDS-PAGE of the expression fractions and IMAC samples of Est16 from *E. coli* BL21 (DE3) cells carrying the pET28a-*est16* vector, which contained the *est16* gene. *E. coli* cells were grown in LB up to  $O.D._{600nm} = 0.5$  at 37 °C and then induced with 0.1 mM IPTG at 28 °C for 20 hours. Lane 1: non-induced cells; Lane 2: induced cells; Lane 3: molecular weight marker (the kDa values are indicated in the picture); Lane 4: soluble extracts; Lane 5: flow-through fraction from the IMAC purification; Lane 6: 10 mM imidazole wash fraction; Lanes 7, 8 and 9: 50 mM, 100 mM and 1 M imidazole elution fractions, respectively. (B) The chromatogram obtained after the size-exclusion chromatography using a HiLoad 16/60 Superdex 200 column (GE Healthcare Bio-Sciences). The insets show the samples related to the peaks loaded into the 12% SDS-PAGE polyacrylamide gels. Lane 1, molecular weight standards; Lane 2: sample before SEC; Lanes 3-4: fractions from the first peak; Lanes 5-12: fractions from the second peak, which were concentrated for further spectroscopic and biochemical analyses.

#### 4.2.4.5 Purified Est16 is active in basic pH and thermostable

After Est16 purification, we performed a preliminary analysis of its activity against tributyrin in Petri dish assays at different pHs. The results clearly showed that Est16 cleaves tributyrin more efficiently in basic pH (8.0 and 9.0) relative to neutral pH (Fig. 5A-C). On the other hand, the loss of activity was not due to differences in the secondary structure of Est16, which reveals an alpha-beta structure, as is expected for esterases/lipases enzymes [37]

(Figure 5D-F). This result is in agreement with the structural model and with the negative peaks of 210 nm (shift from 208 nm) and 222 nm. In accordance with the higher activity at pH 9.0, Est16 showed two steps for complete unfolding (40 °C and 95 °C, respectively), revealing a thermally stable structure (Figure 5G-I).



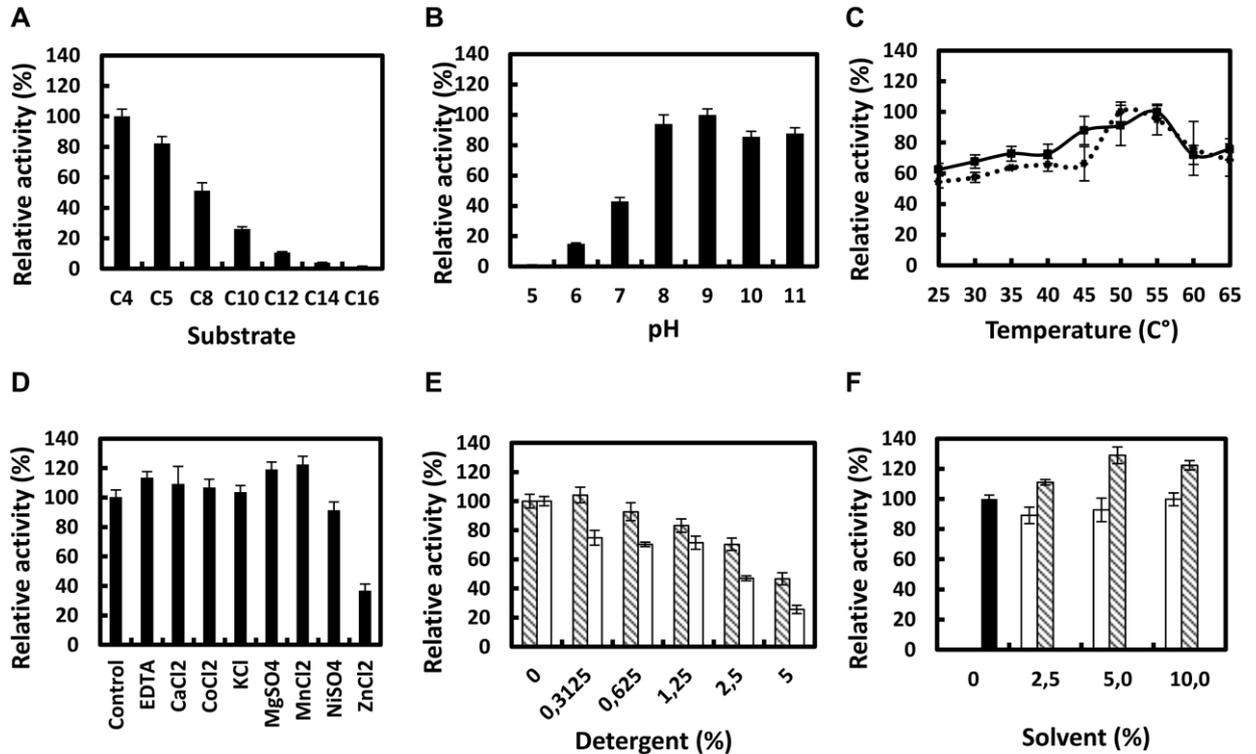
**Figure 5. Tributyrin assay in Petri dishes and spectroscopic analyses of Est16.** Tributyrin assays of Est16 (50  $\mu\text{g}/\text{mL}$ ) in Petri dishes containing 0.23% Tris-HCl (w/v), 1.2% agar (w/v) and 1% tributyrin at pH 7.0 (A), pH 8.0 (B) and pH 9.0 (C). Circular dichroism assays of Est16 at pH 7.0 (D), pH 8.0 (E) and pH 9.0 (F). The samples contained 5.64  $\mu\text{mol}$  of enzyme and the spectra were recorded using a Jasco J-810 spectropolarimeter with a 10 mm path length. Similarly, the thermal denaturation of the enzyme was measured at 222 nm as a function of the temperature at pH 7.0 (G), pH 8.0 (H) and pH 9.0 (I). The melting temperature ( $T_m$ ) is shown for each spectrum.

#### 4.2.4.6 Substrate specificity and kinetic parameters of Est16

The substrate specificity was assessed using *p*-NP esters with different acyl chain lengths ( $\text{C}_4$  to  $\text{C}_{12}$ ). Est16 was able to hydrolyze substrates with acyl chains up to 12 carbons and had high catalytic efficiency toward *p*-NP esters with acyl chain lengths of four and five carbons (Figure 6A). A sharp decrease in the activity was observed for acyl chain lengths greater than 10 carbons. *p*-NP butyrate and valerate were substrates toward which Est16 had higher catalytic efficiencies ( $K_{\text{cat}} \cdot \text{K}_m^{-1}$ ), approximately 60 and 40  $\text{min}^{-1} \cdot \mu\text{M}^{-1}$ , respectively (Table 1). The catalytic efficiency toward butyrate was 4.5, 12.0, and 39.9 times higher than

those of octanoate, decanoate and dodecanoate, respectively. Concerning valerate, the catalytic efficiency was 2.9, 7.9 and 26.6 times higher than those of octanoate, decanoate and dodecanoate, respectively. The preference of Est16 for short chain substrates ( $C < 10$ ) justified its designation as an esterase.

The effect of pH on the enzymatic activity of Est16 was evaluated with *p*NP-butyrate as a substrate. Est16 was found to be an alkaline esterase that is stable in a broad pH range of 7 to 11, but its maximal activity was at pH 8-9 (Figure 6B). Interestingly, Est16 showed activity over the entire temperature range assayed (25 °C to 65 °C) but the optimal temperature range for Est16 activity toward *p*NP-valerate and octanoate was 45 °C to 55 °C (Figure 6C), with maximum activity at 55 °C. This behavior is in agreement with the biophysical studies (Figure 5) and reveals an interesting applicability of this enzyme for biotechnological approaches. The effects of metal ions on the enzyme activity were evaluated by measuring the residual esterase activity in the presence of 0.5 mM of CaCl<sub>2</sub>, CoCl<sub>2</sub>, KCl, MgSO<sub>4</sub>, MnCl<sub>2</sub>, NiSO<sub>4</sub> and ZnCl<sub>2</sub>. The enzyme activity did not show any change in the presence of metals except that Zn<sup>+2</sup> was observed to induce a 40% reduction in activity relative to the control (Figure 6D). Similarly, we tested the effects of detergents and solvents on the activity of Est16. The reaction velocity was measured in the presence of triton X-100, tween 20, DMF and DMSO. Est16 tolerated up to 1.25 % (v/v) of both detergents and showed increased activity with 0.3% of triton X-100 when compared to the control reaction (Figure 6E), suggesting a preference for this detergent. Additionally, the enzymatic catalysis was increased in the presence of the highly polar, water-miscible solvent DMSO and was not affected by up to 10 % (v/v) DMF (Figure 6F). It is possible to speculate that the increased activity in the presence of DMSO was due to stabilization of the enzyme structure and to the increased solubility of the lipophilic substrates.



**Figure 6. Kinetics analyses of Est16.** (A) Substrate specificity against *p*-NP esters of various lengths ( $C_4$  to  $C_{16}$ ); (B) The effect of pH on Est16 activity against *p*-NP butyrate; (C) The effect of temperature on Est16 activity against *p*-NP valerate (■) and octanoate (●); (D) Est16 activity against *p*-NP butyrate was evaluated in the presence of different cations and EDTA; (E) The influence of detergents and (F) organic solvents on the enzyme catalytic activity forward *p*-NP butyrate. In E, the striped bar is triton X-100 and the white bar is tween 20. In F, the black bar is the control, the white bar is DMF and the striped bar is DMSO. All reaction parameters are described in the Materials and Methods section.

**Table 1. Kinetic parameters of Est16 activity against *p*-nitrophenyl substrates.**

Substrate	$K_m$ ( $\mu M$ )	$V_{max}$ ( $\mu M \cdot min^{-1}$ )	$K_{cat}$ ( $min^{-1}$ )	$K_{cat} \cdot K_m^{-1}$ ( $min^{-1} \cdot \mu M^{-1}$ )
<i>p</i> -NP butyrate	$225.1 \pm 31.8$	$206.1 \pm 9.1$	$13211.5 \pm 583.3$	$59.9 \pm 8.5$
<i>p</i> -NP valerate	$320.4 \pm 32.5$	$197.3 \pm 6.8$	$12647.4 \pm 435.9$	$39.9 \pm 4.0$
<i>p</i> -NP octanoate	$628.6 \pm 101.4$	$127.8 \pm 8.6$	$8192.3 \pm 551.3$	$13.4 \pm 2.2$
<i>p</i> -NP decanoate	$1066.0 \pm 116.9$	$82.7 \pm 4.5$	$5301.3 \pm 288.5$	$5.0 \pm 0.6$
<i>p</i> -NP dodecanoate	$969.5 \pm 42.5$	$22.9 \pm 0.5$	$1467.9 \pm 32.1$	$1.5 \pm 0.1$

#### 4.2.5 Conclusions

In this work, a function-based screening of a metagenomic library obtained from oil-contaminated soil was used to search for lipolytic activity. The isolated and characterized enzyme Est16 showed conservation of the sequence and structural features of members from family V of bacterial lipolytic enzymes as well as the patented Anti-Kazlauskas lipase, which shows biotechnological applications in the pharmacological industry. Using structural comparison, spectroscopic and kinetic analyses, we demonstrated that this enzyme has a potential biotechnological application because it shows significant differences in its substrate-binding pocket, it is stable at high temperatures and in the presence of organic solvents, and it also shows a large range of substrates.

#### 4.2.6 Acknowledgments

This work was conducted at the Biochemistry Laboratory of Microorganisms and Plants (LBMP), Department of Technology, São Paulo State University, Jaboticabal Campus, and at the National Laboratory of Biosciences (LNBio), Campinas, Brazil.

#### 4.2.7 References

1. Global Industry Analysts. Industrial enzymes – a global multi-client Market research project. GIA: California; 2004.
2. CBMT.T MaBi. The market and business intelligence company reviews the enzyme market. 2008. Available: <http://www.biomap.fr/biotech/cbdtm-the-market-and-business-intelligence-company-reviews-the-enzyme-market>.
3. Bornscheuer UT, Kazlauskas RJ: Hydrolases in organic synthesis: regio- and stereoselective biotransformations. 1st ed. Weinheim: Wiley-VCH; 2006.
4. Brault G, Shareck F, Hurtubise Y, Lépine F, Doucet N. Short-chain flavor ester synthesis in organic media by an *E. coli* whole-cell biocatalyst expressing a newly characterized heterologous lipase. PLoS One. 2014;9: e91872.

5. Peng Q, Wang X, Shang M, Huang J, Guan G, Li Y, Shi B. Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. *Microb Cell Fact*. 2014;13: 1.
6. Panda T, Gowrishankar BS. Production and applications of esterases. *Appl Microbiol Biotech*. 2005;67: 160-169.
7. Bornscheuer UT. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev*. 2002;26: 73-81.
8. Jin P, Pei X, Du P, Yin X, Xiong X, Wu H, Zhou X, Wang Q. Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. *Biores Tech*. 2012;116: 234-240.
9. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5: 245-249.
10. Rees HC, Grant S, Jones B, Grant WD, Heaphy S. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles*. 2003;7: 415-421.
11. Jungang L, Kegui Z, Wenjun H. Cloning and biochemical characterization of a novel lipolytic gene from activated sludge metagenome, and its gene product. *Microb Cell Fact*. 2010;9: 83.
12. Glogauer A, Martini VP, Faoro H, Couto GH, Müller-Santos M, Monteiro RA, Mitchell DA, Souza EM, Pedrosa FO, Krieger. Identification and characterization of a new true lipase through metagenomic approach. *Microb Cell Fact*. 2011;10: 54.
13. Tao W, Lee M H, Wu J, Kim NH, Lee SW. Isolation and characterization of a family VII esterase derived from alluvial soil metagenomic library. *J Microbiol*. 2011;49: 178-185.
14. Yu EY, Kwon MA, Lee M, Oh JY, Choi JE, Lee JY, Song BK, Hahm DH, Song JK. Isolation and characterization of cold-active family VIII esterases from an arctic soil metagenome. *Appl Microbiol Biotech*. 2011;90: 573-581.
15. Jeon JH, Lee HS, Kim JT, Kim SJ, Choi SH, Kang SG, Lee JH. Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment. *Appl Microbiol Biotech*. 2012;93: 623-631.

16. Seo S, Lee YS, Yoon SH, Kim SJ, Cho JY, Hahn BS, Koo BS, Lee CM. Characterization of a novel cold-active esterase isolated from swamp sediment metagenome. *W J Microbiol Biotech.* 2014;30: 879-886.
17. Roh C, Schmid RD. Isolation of an organic solvent-tolerant lipolytic enzyme from uncultivated microorganism. *Appl Biochem Biotech.* 2013;171: 1750-1758.
18. Fang Z, Li J, Wang Q, Fang W, Peng H, Zhang X, Xiao Y. A novel esterase from a marine metagenomic library exhibiting salt tolerance ability. *J Microbiol Biotech.* 2014;24: 771-780.
19. Steele HL, Jaeger KE, Daniel R, Streit WR. Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotech.* 2009;16: 25-37.
20. Cheeseman JD, Tocilj A, Park S, Schrag JD, Kazlauskas RJ. Structure of an aryl-esterase from *Pseudomonas fluorescens*. *Acta Cryst Sect D Biol Cryst.* 2004;60: 1237-1243.
21. Accorsini FR, Mutton MJR, Lemos EGM, Benicasa M. Biosurfactants production by yeasts using soybean oil and glycerol as low cost substrate. *Braz J Microbiol.* 2012;43: 116-125.
22. Paixão DAA, Dimitrov MR, Pereira RM, Accorsini FR, Vidotti MB, Lemos EGL. Molecular analysis of the bacterial diversity in a specialized consortium for diesel oil. *Rev Bras Ciênc Solo.* 2010;34: 773-781.
23. Wild J, Hradecna Z, Szybalski W. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* 2002;12: 1434-1444.
24. Sambrook J, Russel W. *Molecular Cloning: A Laboratory Manual*. 3rd ed. New York: Cold Spring Harbor Laboratory; 2001.
25. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998;8: 186-194.
26. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 1998;8: 175-185.

27. Astschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman D. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res.* 1997;25: 3389-3402.
28. Arpigny JL, Jaeger KE. Bacterial lipolytic enzymes: classification and properties. *Biochem J.* 1999;343: 177-183.
29. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res.* 1994;22: 4673-4680.
30. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software version 4.0. *Mol Biol Evol.* 2007;24: 1596-1599.
31. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4: 406-425.
32. Casas-Godoy L, Duquesne S, Bordes F, Sandoval G, Marty A. Lipases: an overview. In: Sandoval G, editor. *Lipases and phospholipase.* London: Springer; 2012. pp. 3-30.
33. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234: 779-815.
34. Jiang Y, Morley KL, Schrag JD, Kazlauskas RJ. Different active-site loop orientation in serine hydrolases versus acyltransferases. *Chembiochem.* 2011;12: 768-776.
35. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Cryst Sect D Biol Cryst.* 2004;60: 2126-2132.
36. DeLano, W.L. *The Pymol molecular graphics system, Version 1.2r3pre.* New York: Schrödinger LLC; 2010.
37. Nardini M; Dijkstra BW.  $\alpha/\beta$  Hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol.* 1999;9: 732-737.

#### 4.2.8 Supplementary material

**Table S1 - List of esterases/lipases used for building the phylogenetic tree.** The sequences of members representing the eight families of bacterial lipolytic enzymes were obtained from NCBI (<http://www.ncbi.nlm.nih.gov>) and submitted to the amino acid sequence alignment using ClustalW [29]. The phylogenetic tree was built using MEGA 4.1 [30].

Access number	Microorganism	Protein function	Blast X PDB Query cover/identity (%)
AAC05510.1	<i>Pseudomonas luteola</i>	Triacylglycerol lipase precursor	1OIL 88/93
AAA50466.1	<i>Burkholderia cepacia</i>	Lipase	1OIL 87/94
CAA49812.1	<i>Burkholderia glumae</i>	Lipase	2ES4 89/100
AAC15585.1	<i>Pseudomonas fluorescens</i>	Lipase	3W9U 98/45
AAB01071.1	<i>Proteus vulgaris</i>	Alkaline lipase	3W9U 99/80
CAA32193.1	<i>Pseudomonas fragi</i>	Unnamed protein product	3W9U 95/47
CAA64621.1	<i>Geobacillus thermocatenulatus</i>	Triacylglycerol lipase	2W22 93/99
AAC67547.1	<i>Staphylococcus epidermidis</i>	Lipase precursor	2HIH 64/48
AAA26633.1	<i>Staphylococcus aureus</i>	Lipase precursor	2HIH 59/52
AAB71210.1	<i>Streptomyces cinnamoneus</i>	Lipase LipA	1YS1 29/37
CAA67627.1	<i>Propionibacterium acnes</i>	Triacylglycerol lipase	1T4M 67/26
CAA02196.1	<i>Bacillus pumilus</i>	Lipase	1I6W 100/80
AAA22574.1	<i>Bacillus subtilis</i>	Lipase	1I6W 85/99
AAC38796.1	<i>Salmonella typhimurium</i>	Outer membrane esterase	3KVN 75/28
CAA47020.1	<i>Photobacterium luminescens</i>	Triacylglycerol lipase	3KVN 80/26
CAA37220.1	<i>Moraxella</i> sp.	Unnamed protein product	3VIS 69/46
AAA53485.1	<i>Streptomyces albus</i>	Lipase precursor	1JFR 85/83
AAB51445.1	<i>Streptomyces</i> sp.	Triacylglycerol acylhydrolase	1JFR 84/100
AAC38151.1	<i>Pseudomonas</i> sp.	Lipase	1QZ3 99/45
AAC41424.1	<i>Cupriavidus necator</i>	Lipase-like enzyme	1QZ3 86/37
CAA37862.1	<i>Moraxella</i> sp.	Triacylglycerol lipase	1JJI 76/32
ZP00943646.1	<i>Ralstonia solanacearum</i>	Esterase	3V9A 72/46
AAC67392.1	<i>Sulfolobus acidocaldarius</i>	Lipolytic enzyme	3E3A 83/27
CAA47949.1	<i>Psychrobacter immobilis</i>	Triacylglycerol lipase	4OPM 87/34
CAA37863.1	<i>Moraxella</i> sp.	Triacylglycerol lipase	4OPM

AAC21862.1	<i>Haemophilus influenzae</i>	Esterase/lipase	88/34 3BF7
AAB30793.1	<i>Arthrospira platensis</i>	Serine esterase	88/44 4FHZ
AAC67727.1	<i>Chlamydia trachomatis</i>	Predicted lysophospholipase esterase	97/32 4F21
CAA22794.1	<i>Streptomyces coelicolor</i>	Putative carboxylesterase	84/24 2OGT
Q01470.1	<i>Arthrobacter oxydans</i>	Phenmedipham hydrolase	95/38 2OGT
P37967.2	<i>Bacillus subtilis</i>	Para-nitrobenzyl esterase	95/36 1QE3
AAC60471.2	<i>Pseudomonas fluorescens</i>	Esterase III	100/98 3ZYT
CAA78842.1	<i>Streptomyces anulatus</i>	Esterase A	92/28 3ZYT
AAA99492.1	<i>Arthrobacter globiformis</i>	Carboxylic ester hydrolase	89/34 3ZYT
			99/81

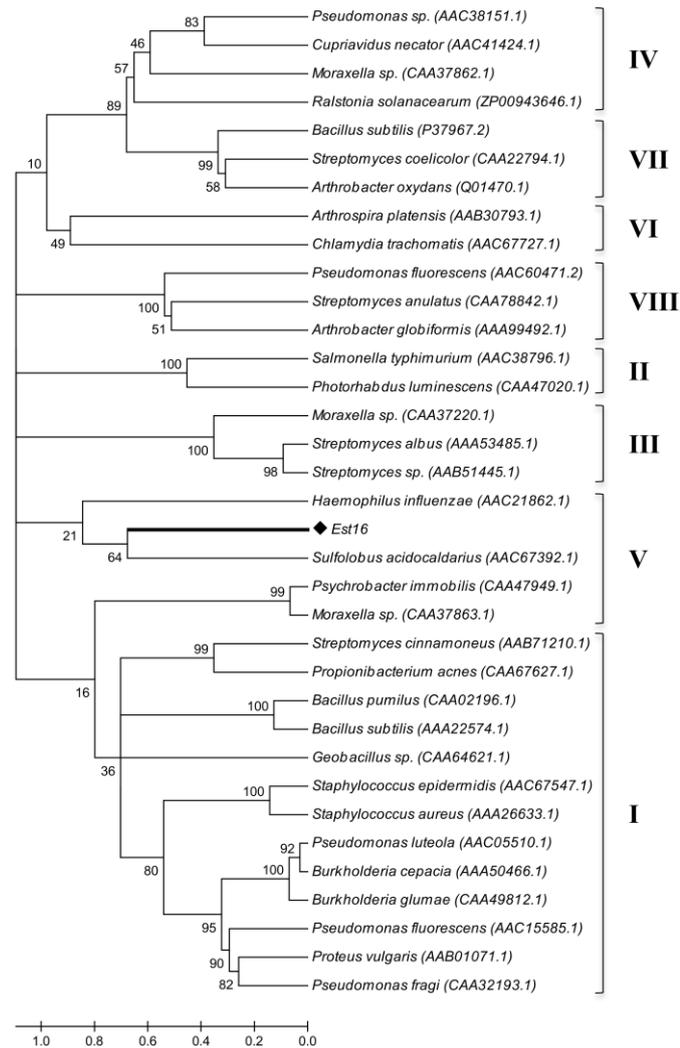
**Table S2 - List of esterases/lipases used for building the phylogenetic tree of Est16 and the patented enzymes.** The patented enzymes were searched for in the literature and the Questel website using the Orbit database. After the patented enzymes were chosen, their sequences were obtained at GenomeQuest website (<http://www.genomequest.com/>). When available, a PDB code for the solved three-dimensional structures of the enzymes is shown.

Reference code	Microorganism	Protein function	Protein identification/ Patent number	Blast X PDB Query cover/ID (%)
Unc_US20050153404_pharma	Unculturable strain	Lipase - preparing of enantiomeric enriched esters and alcohols	US20050153404	1Q0R 89/28
Pfl_P22862_1VA4	<i>Pseudomonas fluorescens</i>	Esterase - increasing the already-existing perhydrolase activity	US7384787	1VA4 99/100
Pal_WO9530744_detergent	<i>Pseudomonas alcaligenes</i>	Lipase - improved resistance to surfactants	WO9530744	1EX9 90/81
Sgr_WO2011150157_detergent	<i>Streptomyces griseus</i>	Lipase - highly suitable as detergent	WO2011150157	3D2C 38/39
Psa_WO2010134035_food	<i>Pelomonas saccharophila</i>	Lipase - combined enzymes, preparing bread and baking	WO2010134035	1GCY 97/94
Sri_WO2010065455_detergent	<i>Streptomyces grimosus</i>	Lipase - removal of oily stains from fabric	WO2010065455	4HYQ 85/65

Ssp_WO2006008653_food	<i>Streptomyces sp.</i>	Lipase - hydrolysis of glycolipids and phospholipids	WO2006008653	-
Lre_US20100216212_pharma	<i>Lactobacillus reuteri</i>	Lipase - anti-obesity activity	US20100216212	-
Tla_WO2006136159_pharma	<i>Thermomyces lanuginosus</i>	Lipase - treatment of human disorders	WO2006136159	2YIJ 32/27
Mgr_WO2007096201_food	<i>Magnaporthe griseae</i>	Lipase - baking, degumming, food emulsifiers	WO2007096201	2YIJ 47/24
Fox_US20030180418_food	<i>Fusarium oxysporum</i>	Lipase - doughs or baked products	US20030180418	2YIJ 44/25
Unk_WO2009106575_food	Unknown strain	Lipase - dairy or baking industry	WO2009106675	2YIJ 45/25
Rsp_WO2007128496_chemicals	<i>Rhodococcus sp.</i>	Esterase - production of fine chemicals	WO2007128496	-
Bsp_WO2008032007_biofuel	<i>Bacillus sp</i>	Lipase - biodiesel production	WO2008032007	-
Dgr_WO2004064537_food	<i>Digitalis grandiflora</i>	Lipase - food emulsifiers	WO2004064537	4LIX 30/28
Unk_WO2007128441_chemicals	Unknown	Esterases - production of fine chemicals	WO2007128441	-
Mth_WO2012078741_industrial	<i>Myceliophthora thermophila</i>	Esterases - industrial food, bioremediation, biorefining, detergents	WO2012078741	24/30
Csp_CN101957838_bioremediation	<i>Culex sp</i>	Esterases - bioremediation of organic phosphorus	CN101957838	4FNM 98/38
Lcu_US5843758_bioremediation	<i>Lucilia cuprina</i>	Esterases - bioremediation of organic phosphate and carbamate	US5843758	4FG5 100/99

---

A



B

Consensus **GXSXGG**  
**Est16** 108 GVKRAHIV**GASMGG**MIAQLV 127 227 REKLK~~N~~LVDV**PAV**VLHGDDP 246 257 AAN**IP**CAELIVV-PGM**GH**D~~F~~ 275  
**Sac AAC67392.1** 134 GYSNLNV**LGS**MGGFVAQQI 153 249 G-QLQNITAP**TLV**IG**G**SD~~L~~ 267 278 AEN**IP**NAQLYIFSPDAG**H**GL 297

**Figure S1. Phylogenetic tree of Est16 with members representing the eight families of bacterial lipolytic enzymes. (A)** A phylogenetic tree of 34 sequences from NCBI and the Est16 protein was generated using the neighbor-joining algorithm (MEGA 4.1 software). The numbers at the nodes indicate the bootstrap percentages from 1,000 replicates. **(B)** The amino acid sequence alignment of Est16 with a member of family V [*Sulfolobus acidocaldarius* (AAC67392.1)]. The residues involved in the substrate-pocket and family classification are shown in bold and the catalytic triad residues are denoted with asterisks (\*).

### 4.3 Chapter 3: Est8 protein

In this chapter, all the results obtained for Est8 that are shown in the manuscript submitted to the Biotechnology and Bioengineering for analysis and publication are

presented. The co-authors' authorization for the inclusion of this chapter is presented in Annex C.

**“High-resolution of Est8 metagenome-derived extends the esterase’s diversity”**

Mariana Rangel Pereira<sup>1,2,3</sup>; Thaís Carvalho Maester<sup>2,3</sup>; Gustavo Fernando Mercaldi<sup>1,4</sup>; Eliana Lemos<sup>3</sup>; Marko Hyvonen<sup>5</sup>; Andrea Balan<sup>6#</sup>

<sup>1</sup>National Laboratory of Biosciences (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo State, Brazil;

<sup>2</sup>University of São Paulo, São Paulo, São Paulo State, Brazil;

<sup>3</sup>Department of Technology, São Paulo State University, Jaboticabal, São Paulo State, Brazil;

<sup>4</sup>Institute of Biology, University of Campinas, Campinas, São Paulo State, Brazil;

<sup>5</sup>Department of Biochemistry, University of Cambridge, United Kingdom; <sup>6#</sup>Department of Microbiology, Biomedical Sciences Institute II, University of São Paulo. Professor Lineu Prestes Avenue, 1374, São Paulo, São Paulo State, Brazil. Post code 05508-900. Email: abalan@usp.br

#Corresponding author

**4.3.1 Abstract**

Esterases catalyze the cleavage and formation of ester bonds and are members of the diverse family of alpha-beta hydrolase fold. They can be useful in industries from different sectors, such as food, detergent, fine-chemicals, biodiesel production, among others. In a previous work, thirty positives clones for lipolytic activity were identified from a metagenomic library of a microbe consortia specializing in diesel oil degradation. Here, a putative gene encoding an esterase/lipase, denominated *est8*, was cloned and the recombinant protein expressed, purified and functional and structurally characterized. The resulting protein, named Est8, was crystallized and had its three-dimensional structure solved at 1.8 Å, which allowed us to characterize the substrate-binding pocket and the location of the conserved catalytic triad Ser157, Asp251 and His281. Est8 is a thermal resistant alkaline esterase with highest catalytic efficiency against *p*-nitrophenyl acetate and stable in the presence of dimethyl sulfoxide. To open the channel for substrate entrance and increase the range of substrates, a double mutant

consisting of mutations M211V and F215G was built and showed improvement in the reaction velocity for *p*-nitrophenyl butyrate and *p*-nitrophenyl valerate, revealing that the use of rational design is an efficient strategy. All data presented in this study reveal that Est8 is an esterase with high biotechnological potential that can be used in the industry for different proposals.

#### **4.3.2 Introduction**

The functional screening tools offered by metagenomics can provide a vast resource to identify uncultivated microorganisms and novel open reading frames, attracting enormous attention of academia and companies (LORENZ; ECK, 2005). Lipolytic enzymes are among biomolecules with potential applicability and can be employed in different sectors such as pure compound production (QUAX; BROEKHUIZEN, 1994), food (CHOI; LEE, 2001; GIULIANI et al., 2001;), paper manufacture (KONTKANEN et al.; 2004) and detergents (JAEGER; REETZ, 1998). These enzymes are widely distributed in animals, plants and microorganisms, and are classified according to their substrate preference as esterase (E.C.3.1.1.1) and lipases (E.C.3.1.1.3), which catalyze the hydrolysis and synthesis of fatty acid esters (ARPIGNY; JAEGER, 1999).

Esterases and lipases are members of a diverse family of alfa/beta hydrolase fold that offers a stable scaffold for the active sites (NARDINI; DIJKSTRA, 1999). According to Ollis and colleagues (1992), enzymes from this family have diverged from a common ancestor preserving the arrangement of the catalytic residues but not the binding site. Esterases can be differentiated from lipases due to the interfacial activation (LONG, 1971), which is observed only in lipases. Lipases present a hydrofobic domain that works as a lid covering the catalytic triad, thus, just in a minimal concentration of substrate the lid moves, exposing the active site (BORNSCHEUER, 2002); wherein, esterases follow the classical Michaelis-Menten kinetics.

Based on biological properties and conserved sequence motifs the bacterial lipolytic enzymes have been classified into eight families (ARPIGNY; JAEGER, 1999) however new families have been proposed over the last years (KIM et al., 2009; LEE et al., 2006; RAO et al., 2013), and it is expected that the number of families increase due to the non-stopping search for new esterases/lipases. Several lipolytic enzymes from family IV has been prospected from different environments through metagenome approach (JEON et al., 2012; JIN et al., 2012; JUNGANG et al., 2010; PENG et al., 2014), however, just few of them were

functional and structurally characterized (BYUN et al., 2007; NAM et al., 2009; NGO et al., 2013). The goal of this study was to overexpress a new esterase metagenome-derived, designated as Est8, for functional and structural characterization. Est8 is a new member of family IV that was isolated from a metagenomic library of a microbe consortium specializing in diesel oil degradation, and shares structural similarities with the highly thermostable esterase from *Pyrobaculum calidifontis* (PDB code 2YH2). We investigated the kinetic, biophysics and structural properties of Est8, being possible to create a double mutant of this enzyme, designated as Est8MF, which displays a larger internal cavity compared to the wild type. Altogether, these data show that Est8 and Est8MF mutant present a set of interesting characteristics, making them potential candidates for biotechnological use.

### **4.3.3 Materials and methods**

#### *4.3.3.1 Screening for lipolytic activity*

As described in our previous work (to be published), a fosmid metagenomic library isolated from a microbe consortium specializing in diesel oil degradation was used for screening potential esterases/lipases. Positive clones were selected based on their ability to hydrolyze tributyrin in Luria-Bertani (LB) agar Petri dishes evidenced by the formation of a clear halo around the colony. From the thirty clones that showed lipolytic activity, PL17.E10 clone was selected for sub-library confection and contig assembly. The gene annotation was carried out and *ORF8* with 936 base pairs was identified encoding Est8 protein.

#### *4.3.3.2 Sequence analysis of est8 gene*

Est8 amino acid sequence was submitted to protein-protein BLAST tool at the National Centre for Biotechnology Information (NCBI, Bethesda, Maryland, USA) for identification of similar proteins and orthologues. Representative sequences of eight families of bacterial lipolytic enzymes, as proposed by Arpigny and Jaeger (1999), were chosen for multiple sequence alignment using ClustalW (THOMPSON et al., 1994) in the BioEdit Sequence Alignment Program (version 7.0.5.3). To add structural information to the tree, we used one representative member of each family to run Blastp against the Protein Data Bank (PDB). The final alignment was used as input for building a phylogenetic tree with Mega 4. 1 (TAMURA et al., 2007) using neighbor-joining algorithm (SAITOU; NEI, 1987). Information

of all sequences used in the alignment and phylogenetic tree are shown in Table S1 (supplementary material).

The DNA sequence of *est8* gene was deposited at GenBank under the accession number of KP699699.

#### 4.3.3.3 DNA amplification and plasmid construction

A DNA fragment of 1047 base pairs containing the *est8* gene (access number KP699699) corresponding to 348 amino acids was amplified by PCR from the positive clone Pl17.E10. The forward 5' CGCGGAATTCACAATGGCACTCGATC 3' and reverse 5' CGGGAAGCTTCCTCACGCCAGCGCC 3' oligonucleotides, were used for the fragment amplification and insertion of the *EcoI* and *HindIII* restriction enzymes sites, respectively in the start and final of the gene. The PCR product was directly cloned into pET28a plasmid (Novagen, Gibbstown, NJ, USA) to generate the pET28a-*est8* vector. The expressed protein from this construct presented a N-terminal His-tagged for further purification, confirmed by DNA sequencing analysis. In order to open the substrate channel in Est8 we also amplified *est8* gene containing two point mutations in the Met211 and Phe215, which were replaced by Val and Gly, respectively. The PCR product amplified using the forward 5' GACGGCGTGATCTGGTTCGGCGATC 3' and reverse 5' GATCGCCGAACCAGATCACGCCGTC 3' oligonucleotides was used as a template to another PCR reaction using the forward 5' TATATCCATGGCACTCGATCCGCAGGCAAAG 3' and reverse 5' TATATAAGCTTACGCCAGCGCCTCTTTCAGCGCCTTC 3' oligonucleotides, which had sites for the restriction enzymes *NcoI* and *HindIII*, respectively (underlined). The mutated sequence was cloned into pHAT2 vector (PERÄNEN et al., 1996), producing pHAT2-*est8mf* vector. The mutated protein codified by pHAT2-*est8mf* was denominated Est8MF.

#### 4.3.3.4 Protein expression, extraction and purification

The recombinant plasmids were transformed into chemically competent *Escherichia coli* BL21 (DE3) cells for overexpression. The transformed cell was grown in LB media containing kanamycin (50 µg mL<sup>-1</sup>) or ampicillin (100 µg mL<sup>-1</sup>) and agitated at 200 rpm at 37 °C until OD<sub>600</sub> reached 0.5-0.6. Est8 and Est8MF proteins were induced, from its respectively pET28a-*est8* and pHAT2-*est8mf* vectors, with 0.4 mM isopropyl β-D-1-thiogalactopyranoside

(IPTG) for 18 hours at 20 °C. After this period, the cultures were centrifuged at 8000 xg for 10 minutes at 4 °C. The cells were resuspended in 30 mL of 50 mM Tris-HCl pH 8.0 and the lysis occurred with Emulsiflex (Avestin Inc, Mannheim, Germany). To the material were added 500 mM NaCl, 10 mM imidazole, 10 mM  $\beta$ -mercaptoethanol, 10% glycerol (v/v). The crude extract obtained was centrifuged at 17.000 xg for 45 minutes at 4 °C to pellet the cell debris. The cell extracts obtained were submitted to the immobilized Ni-NTA affinity chromatography in buffer 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol (v/v). Purified proteins were eluted in the same buffer containing 200 mM imidazole. A second step of purification was performed through size exclusion chromatography to reduce contaminants in a Superdex 200 HiLoad column (GE Healthcare Bio-Sciences, Uppsala, Sweden), at room temperature in buffer 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% glycerol. The purified fractions of Est8 and Est8MF were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions.

#### 4.3.3.5 Circular dichroism and spectroscopic analysis

To evaluate the stability of Est8 and Est8MF in different pH we monitored the secondary structure content and the tryptophan environment in the proteins using, respectively, circular dichroism (CD) and tryptophan intrinsic fluorescence. Far UV CD spectra were recorded on a Jasco-810 spectropolarimeter using a Peltier system PFD 425S for temperature control. Protein samples were prepared at the final concentration of 0.2 mg/mL (5.44  $\mu$ mol and 5.86  $\mu$ mol of Est8 and Est8MF, respectively) in three different pHs: 5 mM sodium phosphate pH 7.0, 5 mM sodium phosphate pH 8.0 and 5 mM CHES pH 9.0. CD spectra were acquired at 10 °C using a 1 mm path length cell at 0.5 nm intervals over the wavelength range from 195 to 260 nm. Ellipticity is reported as the mean residual ellipticity  $[\theta]$  ( $\text{deg.cm}^2.\text{dmol}^{-1}$ ). Samples were also subjected to thermal unfolding from 20 °C to 110 °C with spectra collected at 1 °C intervals. The loss of secondary structure was examined by measuring the ellipticity at 222 nm using 0.5 °C intervals. The  $T_m$  represents the temperature at the midpoint of the unfolding transition. All data collected were corrected for the buffer baseline contribution.

The fluorescence experiment of Est8 was performed at room temperature in K2 multifrequency cross-correlation phase and modulation fluorometer (ISS Inc., Champaign, Illinois, USA). 1  $\mu$ M Est8 was diluted in 5 mM Tris-HCl buffered at pH 7.0, 8.0 and 9.0. Assays using tributyrin as substrate were based on the stoichiometry of 1:2 (protein:ligand)

and samples were incubated for thirty minutes at room temperature before measurements. The tryptophan fluorescence emission was recorded in wavelength range from of 300 to 400 nm.

#### 4.3.3.6 Enzymatic characterization of Est8

The enzymatic activity of Est8 obtained from size-exclusion chromatography was tested in different pH (7.0, 8.0 and 9.0) against tributyrin in Petri dishes containing 0.23% Tris (w/v); agar 1.2% (w/v); 1% tributyrin (v/v). Est8 was prepared at a concentration of 50 µg/ml and 20 µl were added to each dish. The material was incubated at 37 °C for 1 day.

Kinetic parameters of Est8 activity were obtained by measuring the formation of *p*-nitrophenol after *p*-NP acyl esters hydrolysis. The reactions were measured forward an enzyme-free blank to subtract auto-hydrolysis and the released *p*-nitrophenol by spectrophotometrical quantification in 2104 Envision Multilabel Reader (PerkinElmer, Waltham, Massachusetts, USA) at 405 nm ( $\epsilon_{pNP} = 17.000 \text{ M}^{-1} \cdot \text{cm}^{-1}$ ). All assays were carried out in triplicates in 96-well plates with a final volume of 100 µL, using Janus Varispan automated liquid handler (PerkinElmer). Unless otherwise indicated, the reactions were developed at 25 °C using 1 nM of the enzyme, 1 mM of substrates in 50 mM Tris-HCl pH 8.0 and 0.3% triton X-100 (v/v).

The enzyme activity was tested against different *p*-NP acyl esters [acetate (C<sub>2</sub>), butyrate (C<sub>4</sub>) and valerate (C<sub>5</sub>), Sigma, San Louis, Missouri, USA]. The optimum temperature for Est8 activity was evaluated with *p*-NP acetate and the measurements were performed in Enspire™ Plate Reader (PerkinElmer), configured to make a temperature ramp from 25 °C to 65 °C. The pH optimum of Est8 was investigated using buffers of different pH values in order to establish a pH range from 5 to 10 (sodium citrate pH 5.0 and 6.0; Tris-HCl pH 7.0, 8.0 and 9.0; and CHES pH 10.0). These reactions had 0.3% triton X-100 (v/v) and were performed with *p*-NP acetate as substrate.

Est8 was tested for its activity in the presence of metal ions, detergents and solvents against *p*-NP acetate. The influence of metals (ZnCl<sub>2</sub>, NiSO<sub>4</sub>, MnCl<sub>2</sub>, MgSO<sub>4</sub>, KCl, CoCl<sub>2</sub> and CaCl<sub>2</sub>) was measured in the presence of 0.5 mM of each reagents and in the absence of them, as a control. Ultimately, the Est8 tolerance to detergents and solvents was tested in the presence of tween 20 and triton X-100 [0.3125 to 5% (v/v)]; and in the presence of solvents DMF and DMSO [2.5 to 10% (v/v)]. The controls were absent of detergent and solvent, respectively, and were performed in 50 mM Tris-HCl pH 8.0.

The kinetic parameters were performed using 1 nM of Est8 and *p*-NP acyl esters at concentration varying from 0.078 mM to 1 mM at 25 °C. The catalytic rate constant  $k_{\text{cat}}$  ( $\text{mM}^{-1}$ ) was calculated from the initial steady-state velocity according to the equation  $k_{\text{cat}} = V_{\text{max}}/[E]$ .

#### 4.3.3.7 Est8 crystallization

Initial crystallization trials were performed using the following commercial kits: Crystal Screen and SaltRx (Hampton Research, Aliso Viejo, USA), JCSG suite and PACT suite (Qiagen, Hilden, Germany), Precipitant synergy and Wizard I and II (Emerald BioSystems, Bainbridge Island, Washington, USA). First crystals were obtained in 2% polyethylene glycol (PEG) 8000 and 0.8 M triammonium citrate pH 8.5, condition that was refined changing salt, protein concentration and pH parameters. Plates were prepared manually through vapour-diffusion technique using 1  $\mu\text{L}$ :1  $\mu\text{L}$  and 1  $\mu\text{L}$ :0.5  $\mu\text{L}$  rate of protein and mother liquor and then incubated at 18 °C.

#### 4.3.3.8 Data collection, processing and structure refinement

Crystals were submitted to diffraction at the beamline I02 from Diamond Light Source, Oxfordshire, United Kingdom, and images were recorded using a Pilatus 6M detector. Datasets of Est8 were indexed and integrated using XDS (KABSCH, 2010). All structures were solved by molecular replacement using Phaser program in CCP4 package (WINN et al., 2011). The *Alicyclobacillus acidocaldarius* carboxylesterase (PDB code 1EVQ), which shares 43% of amino acid sequence identity with Est8, was used as a model. Refinement of the structures was manually performed using COOT (EMSLEY; COWTAN, 2004) and then subjected to rigid body and restrained refinement cycles using the program REFMAC5 (MURSHUDOV et al., 2011). Protein validation parameters were evaluated according to Ramachandran plot (RAMACHANDRAN et al., 1963). Analysis of the structures and figures were made with Pymol (DELANO, 2010). All statistics from data collection and refinement are shown in Table S2.

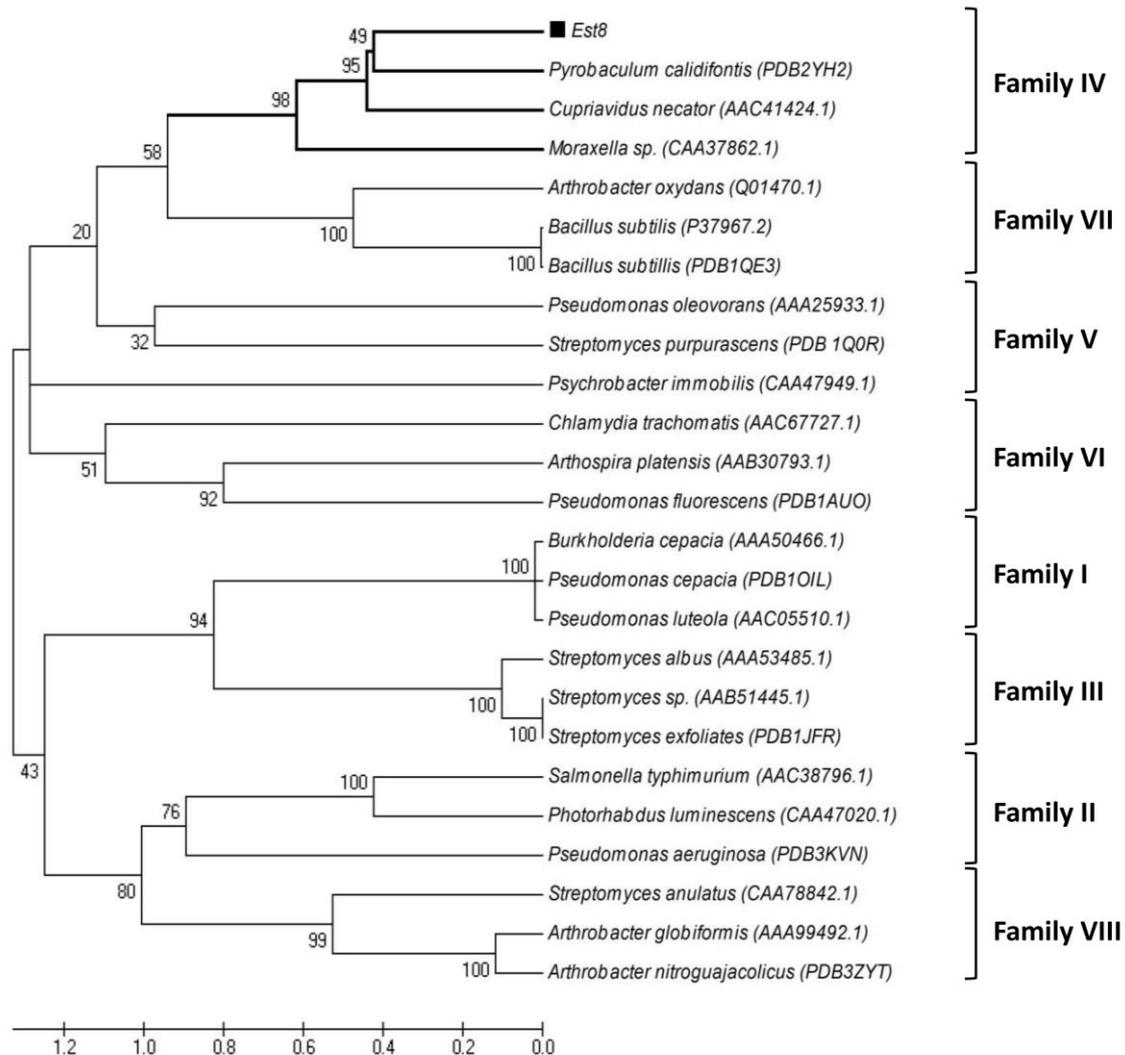
### 4.3.4 Results and discussion

#### 4.3.4.1 Est8 is a family IV member of esterases and lipases

From one contig obtained from a metagenomic library of a microbe consortium specializing in diesel oil degradation we obtained a fragment of 936 bp encoding a putative esterase/lipase, denominated Est8. Its deduced amino acid sequence (311 amino acids) was submitted to Blastp analysis at the NCBI, using the non-redundant protein sequences (nr) database and showed 72% identity with an alpha/beta hydrolase protein from *Parvibaculum lavamentivorans* (accession number WP012110575.1) and 65% identity with a carbohydrate esterase from *Phaeomarinobacter ectocarpi* (accession number CDO59205.1). Moreover, against the metagenomic protein (env\_nr) database Est8 showed 52% identity with hypothetical protein from marine (accession number EDJ20548.1) and freshwater (accession number KGA15039.1) metagenomes, respectively.

To establish the phylogenetic relationships of Est8 with members of the eight families of bacterial lipolytic enzymes as proposed by Arpigny and Jaeger (1999) and esterases/lipases with solved structure a phylogenetic tree was built (Figure 1A). The result evidenced that Est8 is classified in the same branch of two described members from family IV: *Cupriavidus necator* (accession number AAC41424.1) and *Moraxella* sp. (accession number CAA37862.1). In addition, the *Pyrobaculum calidifontis* esterase (PALM et al., 2011; PDB code 2YH2) also was classified in the same family. All the proteins shared conserved sequence motifs as shown in the multiple sequence alignment (Figure 1B), confirming that Est8 belongs to family IV of bacterial lipolytic enzymes, known as hormone-sensitive lipase (HSL) family. Est8 contains the conserved residues from the catalytic triad <Ser159-Asp253-His283> which the catalytic nucleophile Ser is positioned in the consensus pentapeptide GDSAG. Additionally, Est8 presents a strictly conserved HGGG motif, <H85-G86-G87-H88>, located upstream of the active site, whose function is related to formation of hydrogen bonds to stabilize the oxyanion hole during catalysis (Wei et al., 1999).

A



B

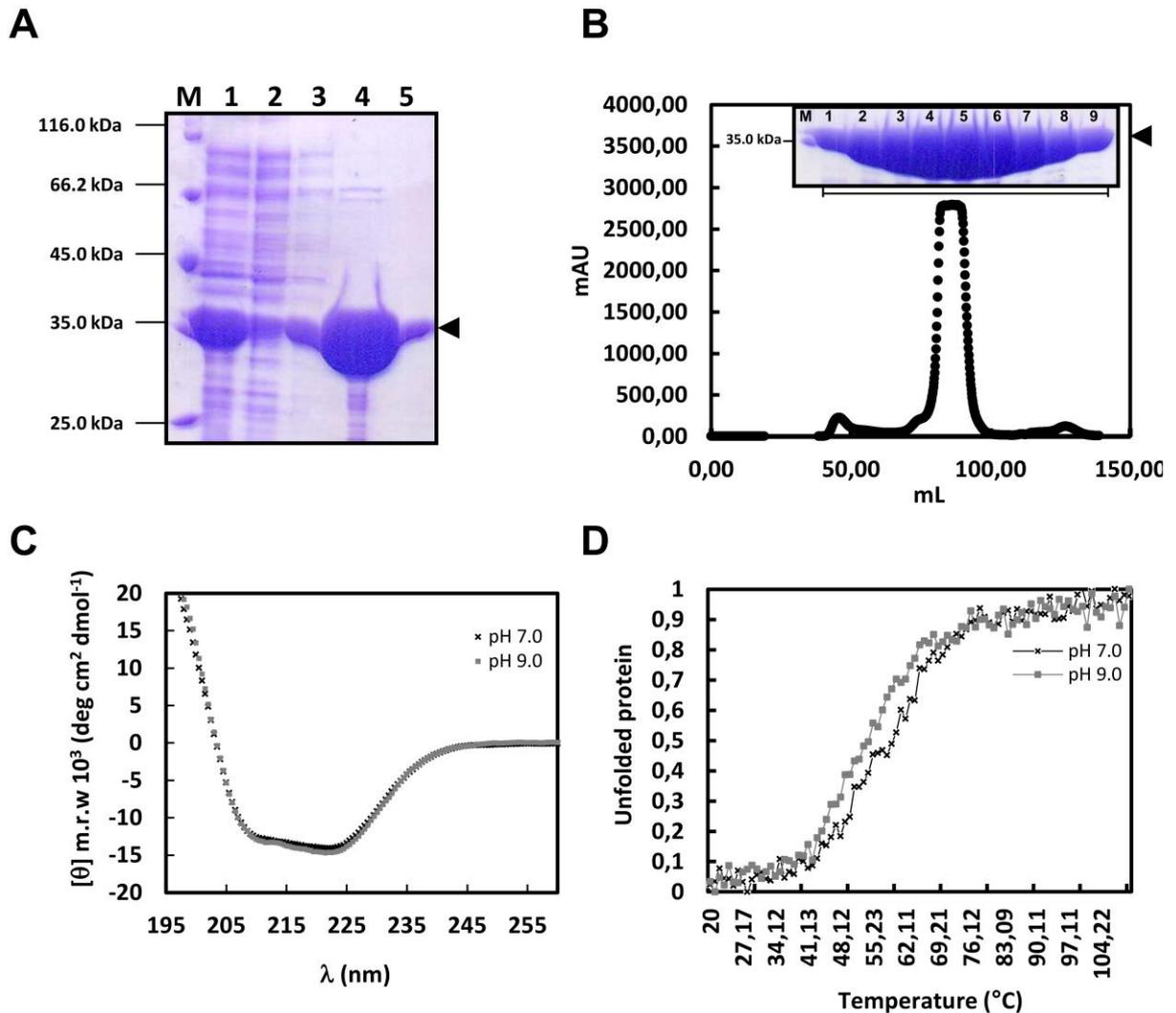
EST8	81	LVYFHGGGWV	90	155	VAGDSAGGNLA	164	251	AGFDPLKDEGRAYAEALKAAGVPTEYVNYEGMIHGF	285
AAC41424.1	126	LVYFHGGGFT	135	200	VGDSAGGTLA	209	300	AGYDPLHDEGVAYAEKLRAGVAATLADYPGMIHDF	334
CAA37862.1	161	MLFFHGGGFC	170	235	LSGDSAGGCCLA	244	361	AELDILRDEGLAYAEELLQKEGVQVQTYTVLGAHPGF	393
PDB: 2YH2	79	VVYFHGGGFV	88	153	VAGDSAGGNLA	162	253	AEYDPLRDEGELYAHLKTRGVRVAVRYNGVIHGF	286
		:::****:			:.***** **			* * *:*** ** .: ** . * **	

**Figure 1. Phylogenetic analysis of Est8 and structural characteristics.** (A) Phylogenetic tree of Est8 protein and representative members of bacterial lipolytic enzymes. All sequences were extracted from NCBI and PDB to represent the 8 lipolytic families. Multiple sequence alignment was performed with ClustalW and used as input for building the phylogenetic tree using Mega 4.1 program and neighbor-joining method. The numbers at the nodes indicate the bootstrap percentages of 1000 replicates. Scale bar: number of amino acid substitutions. (B) Sequence characteristics of the substrate-binding sites from proteins belonging to family IV and a three-dimensional structure available at PDB. Multiple sequence alignment of the partial amino acid sequences containing the conserved blocks of Est8, two members of family IV [*Cupriavidus necator* (AAC41424.1), *Moraxella* sp. (CAA37862.1)] and an esterase (PDB code 2YH2) from hyperthermophilic microorganism *Pyrobaculum calidifontis*. Amino acid residues belonging to the catalytic triad are pointed by an arrow (▼) and motifs from family IV are colored.

#### 4.3.4.2 Recombinant Est8 was expressed as a soluble, folded and stable protein

The *est8* gene was cloned at the N-terminal end in pET28a expression vector to overexpress the recombinant protein for biochemical and spectroscopy characterization and crystallization assays. The recombinant Est8 protein was expressed in *E. coli* BL21 (DE3) strain and the molecular mass was 37 kDa, as expected from its amino acid sequence. The target protein was firstly purified by Ni-NTA affinity chromatography from 10 mM imidazole buffer (Figure 2A), and then submitted to a second purification step using size exclusion chromatography (Figure 2B). Two peaks were obtained in this purification but Est8 recombinant protein was recovered from second peak with a final yield around 110 mg/mL per liter of induced culture and remained stable even at high concentrations (10-20 mg/mL) or prolonged storage at 4 °C what was an important step for spectroscopic assays and crystallization trials.

Purified Est8 was submitted to circular dichroism analysis in neutral (7.0) and basic (9.0) pH. The spectra obtained showed a typical profile of an alpha-beta protein with negative peaks at 208 nm (shift from 208 nm) and 222 nm, which is in accordance for esterases/lipases (NARDINI; DIJKSTRA, 1999). Moreover, the spectra revealed that the pH changes did not alter the Est8 secondary structure (Figure 2C) or the denaturing temperature (Figure 2D). Est8 showed to be a thermal stable protein, since the  $T_m$  obtained in both pHs assayed was 55 °C (Figure 2D).



**Figure 2. Expression, purification and biophysical analysis of Est8.** (A) *E. coli* BL21(DE3) cells carrying the pET28a-*orf8* vector were induced with 0.4 mM IPTG for 18 hours at 20 °C and their cellular extracts were submitted to IMAC purification using nickel-bound columns as described in Materials and Methods session. M: molecular weight standard (kDa values indicated in the picture); 1: soluble extract; 2: flow-through fraction; 3-5: fraction eluted with buffer containing 10 mM, 200 mM and 500 mM imidazole, respectively. (B) Chromatogram obtained in the size-exclusion chromatography of Est8 using a HiLoad 16/60 Superdex 200 column (GE Healthcare Bio-Sciences). Fractions obtained in the large peak are shown as an inset in a SDS-12% polyacrylamide gel. M: molecular weight standard; 1-9: eluted fractions representing the larger peak. (C) Circular dichroism of 5.44  $\mu$ mol Est8 at pH 7.0 (x) and pH 9.0 (■). Spectra were recorded using Jasco J-810 spectropolarimeter. (D) Thermal denaturation of Est8 was measured at 222 nm as function of the temperature at pH 7.0 (x) and pH 9.0 (■), showed in gray and black lines, respectively.

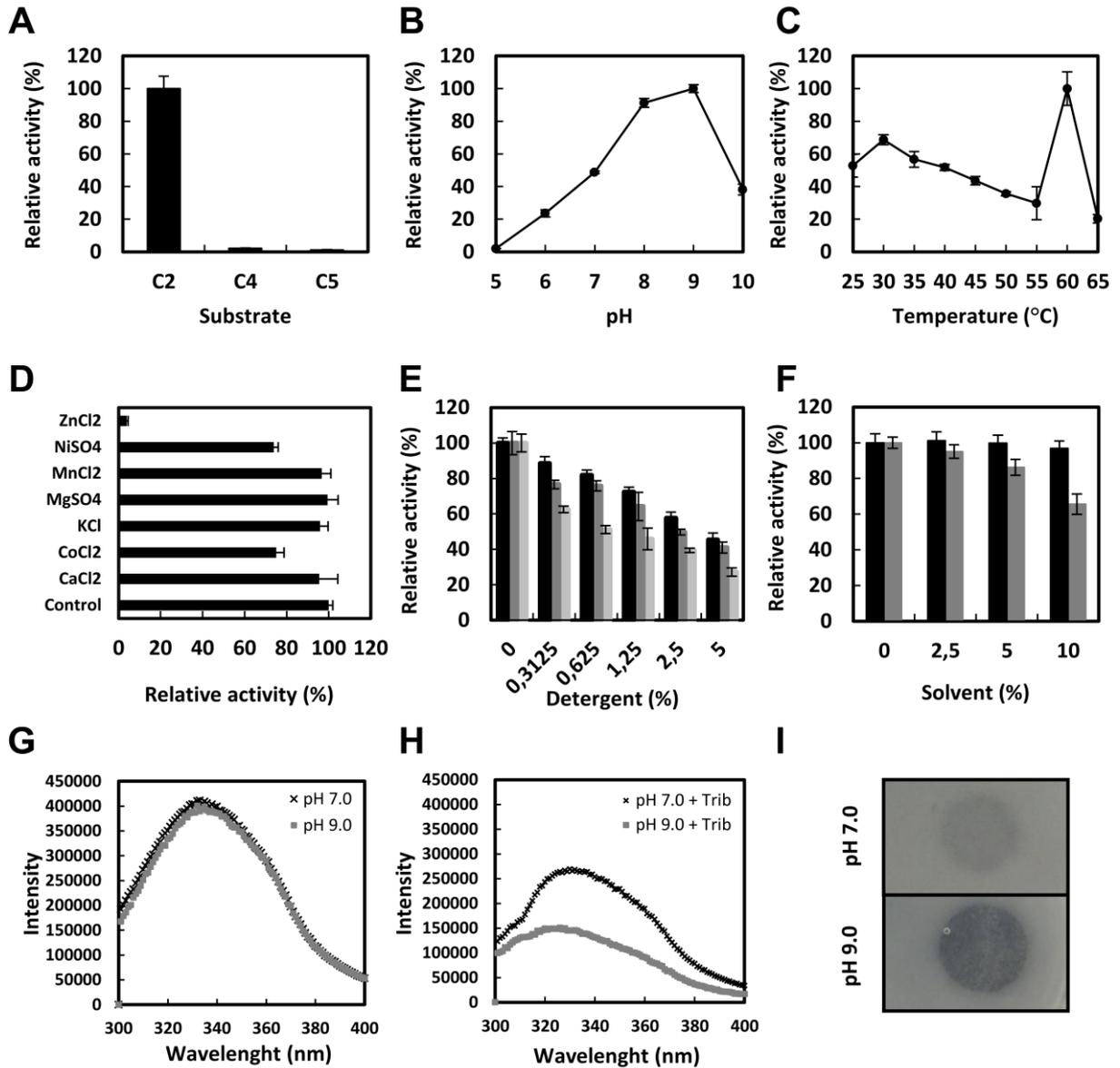
#### 4.3.4.3 Kinetics characterization of Est8

Kinetics analysis of Est8 were carried out to determine the best conditions for enzymatic activity and substrate range. A substrate spectrum was recorded with *p*-nitrophenyl (*p*-NP). Est8 was able to hydrolyze acyl chain-lengths up to five carbons, with highest activity toward *p*-NP acetate (C<sub>2</sub>), whereas, with *p*-NP butyrate (C<sub>4</sub>) and *p*-NP valerate (C<sub>5</sub>) is noticed a decrease in activity (Figure 3A). The catalytic efficiency ( $K_{cat} \cdot K_m^{-1}$ ) against *p*-NP acetate was 397.14 min<sup>-1</sup>·μM<sup>-1</sup> (Table 1), followed by 25.47 and 3.34 min<sup>-1</sup>·μM<sup>-1</sup> toward butyrate and *p*-NP valerate, respectively. It was verified that catalytic efficiency of Est8 against *p*-NP acetate was 15.6 and 118.9 times higher than *p*-NP butyrate and *p*-NP valerate, respectively. The kinetic study of Est8 shows its preference for short acyl-chain lengths (C<10), that allowed its denomination as esterase.

Using 1 mM of *p*-NP acetate as a substrate, the Est8 protein showed to be an alkaline esterase with maximum activity at pH 9.0 when tested in 50 mM Tris-HCl (Figure 3B). Above this pH, the activity was rapidly decreased. Analysing the temperature effect on Est8 with *p*-NP acetate as substrate, this enzyme revealed optimum temperature of 60 °C (Figure 3C) and, interestingly, for the entire temperature range measured (25 °C to 65 °C) the relative activity is above 30%. This profile is consistent with data obtained in biophysical studies (Figure 2), that reveals the melting temperature around 55 °C showing its applicability in biotechnology as a thermal enzyme.

In the presence of metal ions the activity of Est8 showed no significant variation compared to the control, except for ZnCl<sub>2</sub>, NiSO<sub>4</sub> and CoCl<sub>2</sub>, which reduced the activity by 3.9, 73.9 and 74.9, respectively (Figure 3D). In view of testing the effect of detergents and solvents on Est8 activity, it was developed assays in the presence of triton X-100, NP-40, tween 20, DMF and DMSO. Est8 displayed an activity reduction for all detergents (Figure 3E); however, up to 5% of these detergents it is still possible to verify 27% of activity. Also, comparing the Est8 activity with detergents to control is verified that Est8 had preference for triton X-100. Impressively, the activity of Est8 did not decrease significantly in the presence of up to 10% of organic solvents (Figure 3F). For example, the activity of Est8 with 10% DMSO had a reduction of 3.2% when compared to the control. Verifying Est8 activity in the presence of these solvents, it is noticed the higher activity with DMSO compared to DMF, probably, due to this high polar water-miscible solvent favor the increase of substrate solubility.

We also observed that pH had no effect on the tryptophan fluorescence in absence of substrate (Figure 3G). However, addition of trybutirin induced a significant quenching of the tryptophans fluorescence (Figure 3H), which is in accordance with the presence of two tryptophans (W87 and W213) in the active site. At pH 9.0 we observed a more accentuated quenching as well as the displacement of the maximum peak of the spectrum to 320 nm suggesting the tryptophans are located in a more hydrophobic region. Together with the Est8 activity against trybutirin in Petri dishes plates, that evidenced a higher activity of the protein in basic pH (Figures 3I), these data suggested that Est8 assumes a more efficient conformation for interaction with the substrate and activity at pH 9.0.



**Figure 3. Enzymatic characterization of Est8.** (A) Substrate specificity against *p*-nitrophenyl acetate (C<sub>2</sub>), *p*-nitrophenyl butyrate (C<sub>4</sub>) and *p*-nitrophenyl valerate (C<sub>5</sub>). From B to F, the enzyme activity was against *p*-nitrophenyl acetate (C<sub>2</sub>) using 1 nM of protein. (B) Effect of pH in Est8 activity. Reaction of Est8 against C<sub>2</sub> was carried out varying the pH range. (C) Effect of temperature. The measures were undertaken in a temperature range from 25 °C to 65 °C; (D) Est8 activity in presence of different cations. Control sample was prepared with no cations; (E) Influence of detergents. Black bar: triton X-100, dark gray: NP-40, and gray: tween 20. (F) Organic solvents. Black bar: DMSO and dark gray: DMF. Data are mean values of three independent measures and the bars indicate the corresponding standard deviation. Analysis of tryptophan intrinsic fluorescence of Est8 is shown in absence (G) and presence of tributyrin (Trib) (H) at pH 7.0 (x) and pH 9.0 (■). Protein was used at concentration of 1 μM, calculated according to the extinction coefficient, and tributyrin at 2 μM with measurements made at room temperature. (I) Enzymatic assay of Est8 against tributyrin in Petri dishes plates, evaluated by the halo presence.

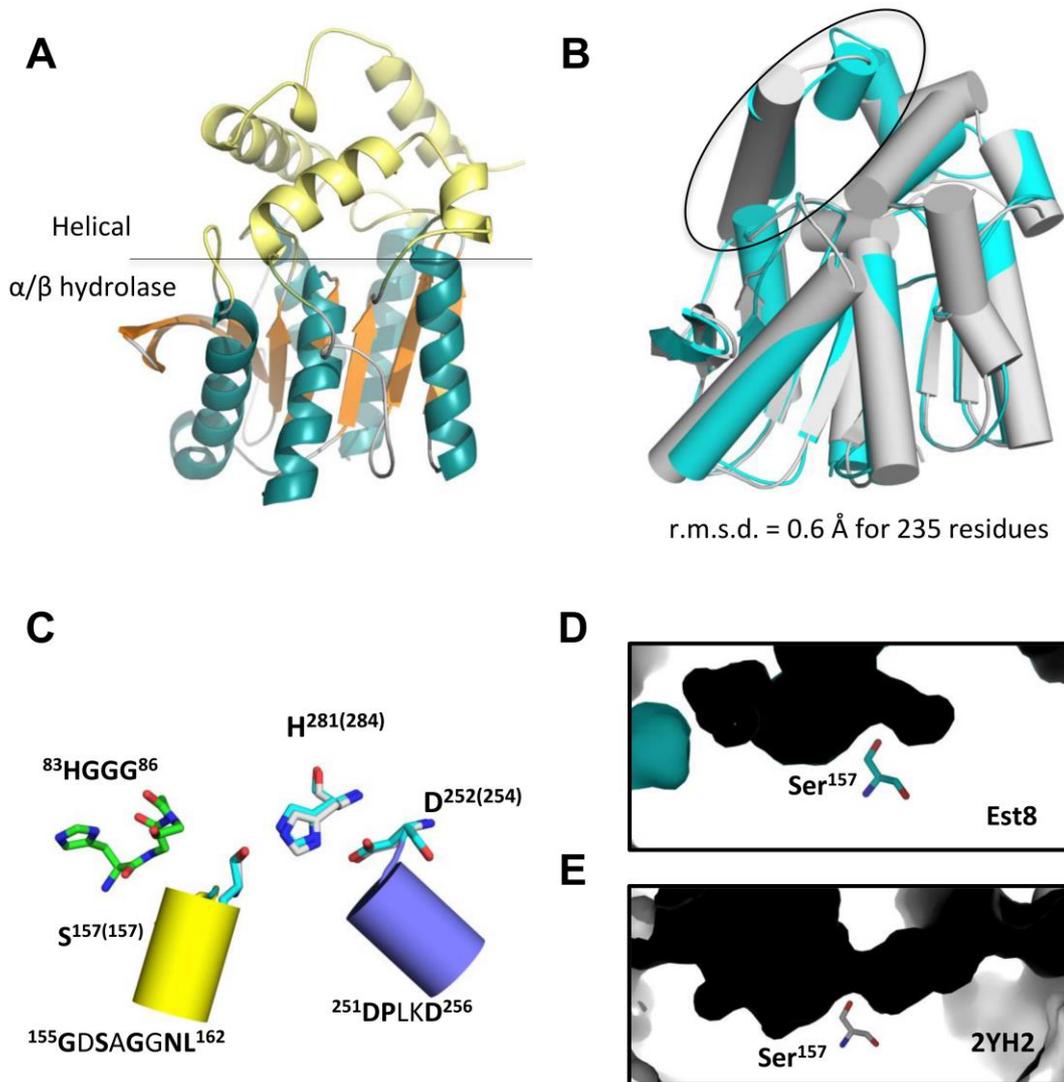
#### 4.3.4.4 Crystallization and structure determination

Est8 crystal growth occurred until 6 days leading to bi-pyramidal crystals with dimensions ranging from 100 to 500  $\mu\text{m}$ . Crystals were obtained in different crystallization conditions containing PEG, MES or HEPES in a pH range of 6.5 to 7.5. Refinement of all crystallization conditions led to crystals grown in 1% PEG 8000, 1.4 M Triammonium citrate pH 9.0 using 17.8 mg/mL of pure protein, which diffracted at maximum of 1.85  $\text{\AA}$  resolution. Crystals of Est8 showed the symmetry and systematic absences of the primitive hexagonal space group  $P3_12_1$ . Table S2 summarizes the data-collection statistics. The Matthews coefficient (MATTHEWS, 1968) was 3.35  $\text{\AA}^3/\text{Da}$  and the solvent content 63.4% corresponding to 2 molecules in the asymmetric unit. After structure refinement the statistics showed  $R_{\text{factor}}/R_{\text{free}}$  of 17.1%/18.91%, respectively (resolution range 30.0 - 1.85  $\text{\AA}$ ) (Table S2). The final structure of Est8 displayed the conserved structure of esterases and lipases that shows an ellipsoidal shape defined by an alpha-beta hydrolase fold, in which two domains are evident: one for sustaining ( $\alpha/\beta$  domain) and the proper domain that forms the active site (helical domain) (Figure 4A, green/orange and pale yellow, respectively). Alpha-beta domain has a mixed  $\beta$ -sheet composed of 8  $\beta$ -strands (only strand  $\beta_2$  is antiparallel) surrounded by ten helices: Pro3-Ala13, Ile20-Asp22, Val25-Leu39, Thr94-Ala106, Ala125-Ile144, Ala158-Ala173, Ala195-Asn200, Arg208-Leu219, Leu253-267, Asp291-Ala308. A small cap domain composed of Pro3-Pro24 and Arg208-Leu219 (Figure 4B, circle) is located on top of the helical domain probably controlling the substrate entry and access to the active binding pocket.

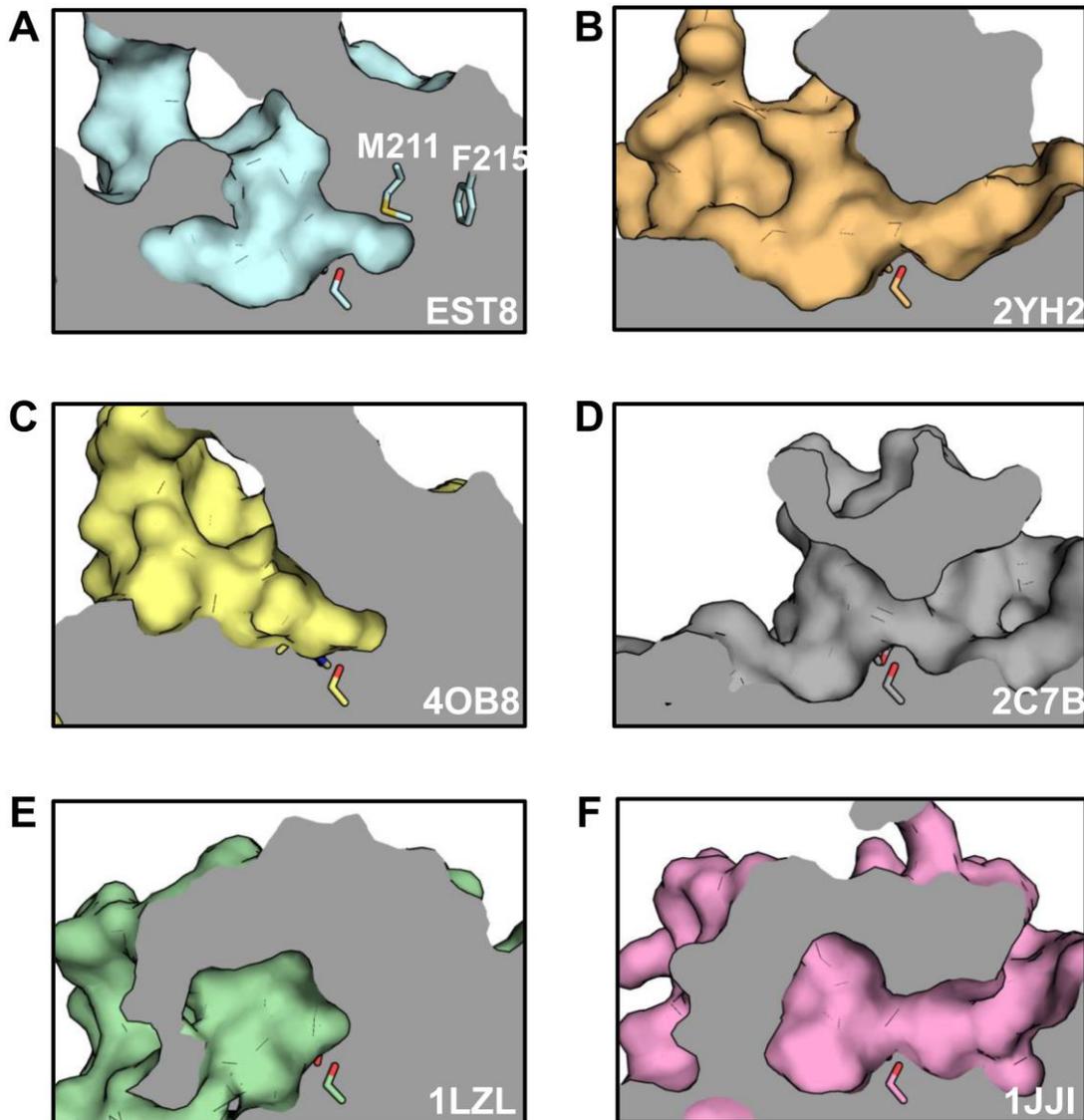
In order to compare Est8 structure with the *Pyrobaculum calidifontis* esterase we carried out a structural superposition that showed a highly conserved  $\alpha/\beta$ -fold (root mean square deviation, r.m.s.d. of 0.6  $\text{\AA}$ ) (Figure 4B) and the catalytic triad residues (Figure 4C, cyan stick). The catalytic nucleophile Ser157 is positioned between  $\beta_5$  and  $\alpha_6$ , in the consensus pentapeptide GDSAG (yellow cartoon), while the other two residues (Asp251 and His281) were found between  $\beta_7$  and  $\alpha_9$  and between  $\beta_8$  and  $\alpha_{10}$ , respectively (Figure 4C, cyan stick). The three amino acids in the catalytic triad were located in close proximity with the residues that are involved in the formation of oxyanion hole during catalysis (WEI et al., 1999) such as Gly84-86, located in the strictly conserved HGGG motif (Figure 4C, green stick).

In addition, we identified two additional motifs that also are involved on catalysis in members of family IV (HSL-family), such as  $^{155}\text{GXSXGGNL}^{162}$  (yellow cartoon), as mentioned before, and  $^{251}\text{DPXXD}^{256}$  (blue cartoon). These data corroborated the previous

phylogenetic analysis that Est8 is a member of this family sharing the same catalytic mechanism. Interestingly, although both proteins belong to the same family and conserve exactly the same three-dimensional structure, when the ligand binding pockets are compared, we observed quite different channels (Figure 4D and E, in black). Moreover, a Blastp using Est8 sequence versus the Protein Data Bank returned at least 5 structures of esterases/lipases with amino acid sequence identity higher than 45%. Again, structurally, all proteins shared the same fold with slight differences regarding the helical domain. Despite that, the channels for substrate were quite different when compared (Figure 5, showed in surface colored). These differences that change the cavity and electrostatic potential, are important for differentiation, selection and cleavage of the substrates. Interestingly, we observed that the two residues Met211 and Phe215 blocked the channel in Est8. To verify if the opening of the channel would induce differences in the activity of the enzyme, we built a double mutant Met211Gly/Phe215Val (Est8MF).



**Figure 4. Overall structure of Est8.** (A) Cartoon representation of Est8 three-dimensional structure evidencing the helical (yellow) and  $\alpha/\beta$  (green and orange) domains. The  $\alpha/\beta$  domain consists of a  $\beta$ -sheet (orange) with 8 strands surrounded by helices (green) forming a structure that sustains the small and helical domain where the active site is formed. (B) Structural superposition of Est8 and *Pyrobaculum calidifontis* esterase (PDB code 2YH2). The presence of a cap domain in the helical domain is evidenced in a black circle. This structural feature is related to the accessibility to the active site. R.m.s.d. between both proteins is shown. (C) Overall positioning and detail of Est8 substrate-binding site (cyan stick) superimposed with *P. calidifontis* esterase (gray stick). Residues from the HGGG motif are shown in green stick. Yellow and blue cartoon show the two additional motifs that describe family IV of bacterial lipolytic enzymes. Detailed comparison between Est8 (D) and *P. calidifontis* esterase (E). The nucleophilic Ser157 is shown in both structures.

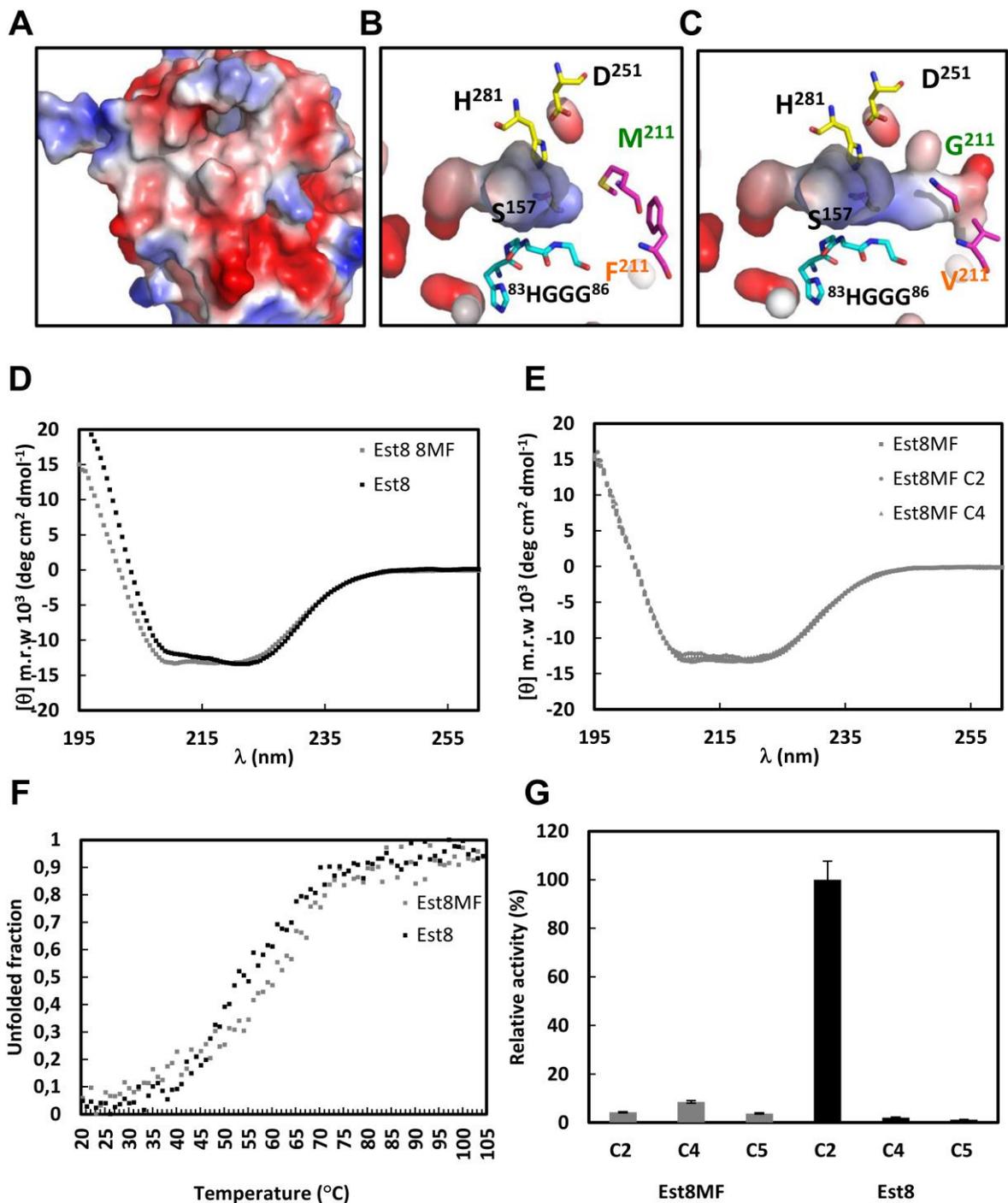


**Figure 5. Comparison of the substrate channel in different esterases/lipases that share similar fold with Est8.** Proteins are shown as surface and channels are colored. The enzymes are (A) Est8; (B) *Pyrobaculum calidifontis* esterase (PDB code 2YH2); (C) *Pseudomonas putida* ECU1011 esterase (PDB code 4OB8); (D) ESTE1 carboxiesterase (PDB code 2C7B); (E) heroin esterase (PDB code 1LZL); (F) *Archaeoglobus fulgidus* carboxiesterase (PDB code 1JJI).

#### 4.3.4.5 The opening of the channel for substrate in Est8 increased its activity against C4 and C5

Est8MF mutant was built based on rational analysis of the activity site and had two residues mutated: Met211 to Gly for drastically reduction of the side chains and chemical properties, and Phe215 to Val for reducing the side chain but maintaining the hydrophobic profile. Based on the structure of Est8, a mutated protein was built after replacement of the

residues and channels were compared (Figure 6A to C). As expected, the short channel in Est8 is changed for a long one in the mutant protein. The expressed protein was obtained after induction of *E. coli* cells carrying the vector pHAT2 as described in Materials and Methods section. Est8MF showed soluble and folded conformation when compared to the wild type protein revealing slight differences in the helix content, as observed in the region 195 to 210 nm from the spectra and in presence of substrates as C2 and C4 (Figure 6D and E). The mutations also induced small changes in the melting temperature of the mutated protein that was increased in 3 °C compared to the wild type (Figure 6F). Finally, the enzymatic activity of the mutant protein was compared with the wild type in presence of the substrates (Figure 6G). As expected, the mutation induced an increasing of 4 and 2 fold in the activity of enzyme against the substrates C4 and C5, respectively, indicating that the opening of the channel is one of the important features for selection of substrates and activity (Table 1).



**Figure 6. Structural and functional characteristics of Est8MF.** (A) The electrostatic potential of Est8 is shown as surface; (B) Est8 channel evidencing the residues Met211 and Phe215 that block the substrate entrance; (C) Model of the opened channel in Est8MF; (D) Circular dichroism spectra of Est8 and its mutant at pH 8.0 (E) and in presence of substrates C<sub>2</sub> and C<sub>4</sub>; (F) Thermal denaturation of Est8; (G) Enzymatic activity of Est8MF compared to the wild type in presence of *p*-nitrophenyl acetate (C<sub>2</sub>), *p*-nitrophenyl butyrate (C<sub>4</sub>) and *p*-nitrophenyl valerate (C<sub>5</sub>).

**Table 1. Kinetic parameters of Est8 and Est8MF against different substrates.**

Protein	T(°C)	Substrate	Vmax ( $\mu\text{M}\cdot\text{min}^{-1}$ )	Kcat ( $\text{m}^{-1}$ )	Km ( $\mu\text{M}$ )	Kcat/Km ( $\text{min}^{-1}\cdot\mu\text{M}^{-1}$ )
Est8	25	<i>p</i> -NP acetate (C <sub>2</sub> )	89.0 ± 3.15	89040 ± 3150	224.2 ± 21.1	397.1 ± 148.9
	25	<i>p</i> -NP butyrate (C <sub>4</sub> )	1.4 ± 0.07	1423 ± 70	55.8 ± 9.8	25.5 ± 7.1 <sup>1</sup>
	25	<i>p</i> -NP valerate (C <sub>5</sub> )	0.82 ± 0.03	820 ± 31	245.1 ± 29.6	3.3 ± 1.0
Est8MF	25	<i>p</i> -NP acetate (C <sub>2</sub> )	0.68 ± 0.10	680 ± 100	124.9 ± 62.60	5.4 ± 1.6
	25	<i>p</i> -NP butyrate (C <sub>4</sub> )	3.94 ± 0.13	3939 ± 130	745.4 ± 60.64	5.3 ± 2.1
	25	<i>p</i> -NP valerate (C <sub>5</sub> )	1.51 ± 0.25	1510 ± 250	780.0 ± 246.7	1.9 ± 1.0

#### 4.3.5 Conclusion

In this work we showed the functional and structural characterization of an esterase isolated from metagenomic approach. The enzyme is a new member of family IV of bacterial lipolytic enzymes, thermal stable and resistant to different additives. The best substrates were C2 and C4, but comparing to solved three-dimensional structures of similar enzymes, we showed that despite the high similarity in the fold, the active site can show variations even in members of the same family. Based on the solved structure of Est8 we built a mutant protein with longer channel that showed increased activity against substrates of long chain length. Altogether, these data detach the importance of the methodology for discovery of new enzymes and improvement of their activities for biotechnological proposals.

#### 4.3.6 Acknowledgments

This work was supported by São Paulo State Foundation (FAPESP), grant numbers 2011/09136-7 and 2012/20490-0 (Research Internships Abroad - BEPE). We thank the infrastructure offered by the São Paulo State University, Jaboticabal Campus, Brazil; the National Laboratory of Biosciences (LNBio), Campinas, Brazil; and the Department of Biochemistry, University of Cambridge.

#### 4.3.7 References

ARPIGNY, J. L.; JAEGER, K-E. Bacterial lipolytic enzymes: classification and properties. **Biochemical Journal**, v. 343, p. 177-183, 1999.

BYUN, J-S.; RHEE, J-K.; KIM, N. D.; YOON, J.; KIM, D-U.; KOH, E.; OH, J-W.; CHO, H-S. Crystal structure of hyperthermophilic esterase EstEI and the relationship between its dimerization and thermostability properties. **BMC Structural Biology**, v. 7, 2007.

BORNSCHEUER, U. T. Microbial carboxyl esterases: classification, properties and application in biocatalysis. **FEMS Microbiology Reviews**, v. 26, p. 73-81, 2002.

CHOI, Y. J.; LEE, B. H. Culture conditions for the production of esterase from *Lactobacillus casei* CL 96. **Bioprocess and Biosystems Engineering**, v. 24, p. 59-63, 2001.

EMSLEY, P.; COWTAN, K. Coot: model-building tools for molecular graphics. **Acta Crystallographica Section D, Biological Crystallography**, p. 2126-2132, 2004.

DELANO, W.L. The Pymol molecular graphics system, Version 1.2r3pre. New York: Schrödinger LLC; 2010.

GIULIANI, S.; PIANA, C.; SETTI, L.; HOCHKOEPLER, A.; PIFFERI, P.G.; WILLIAMSON, G.; FAULDS, C.B. Synthesis of pentylferulate by a feruloyl esterase from *Aspergillus niger* using water-in-oil microemulsions. **Biotechnology Letters**, v. 23, p. 325-330, 2001.

JAEGER, K., REETZ, M. T. Microbial lipases form versatile tools for biotechnology. **Trends in Biotechnology**, v. 16, p. 396-403, 1998.

JEON, J. H.; LEE, H. S.; KIM, J. T.; KIM, S. J.; CHOI, S. H.; KANG, S. G.; LEE, J. H. Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment. **Applied Microbiology and Biotechnology**, v.93, p. 623-631, 2012.

JIN, P.; PEI, X.; DU, P.; YIN, X.; XIONG, X.; WU, H.; ZHOU, X.; WANG, Q. Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. **Bioresource Technology**, v. 116, p. 234-240, 2012.

JUNGANG, L.; KEGUI, Z.; WENJUN, H. Cloning and biochemical characterization of a novel lipolytic gene from a activated sludge metagenome, and its gene product. **Microbial Cell Factories**, v. 9, 2010.

KABSCH, W. XDS. **Acta Crystallographica Section D, Biological Crystallography**, p. 125-132, 2010.

KIM, E-Y.; OH, K-H.; LEE, M-H.; KANG, C-H.; OH, T-K.; YOON, J-H. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. **Applied and environmental microbiology**, p. 257-260, 2009.

KONTKANEN, H.; TENKANEN, M.; FAGERSTROM, R.; REINIKAINEN, T. Characterisation of steryl esterase activities in commercial lipase preparations. **Journal of Biotechnology**, v. 108, p. 51-59, 2004.

LEE, M. H.; LEE, C. H.; OH, T. K.; SONG, J. K.; YOON, J. H. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: Evidence for a new family of bacterial lipases. **Applied and Environmental Microbiology**, v. 72, p. 7406-7409, 2006.

[LONG, C. **Biochemists handbook**. London: Redwood, 1971. 1192?p.]

LORENZ, P.; ECK, J. Metagenomics and industrial applications. **Nature**, v. 3, p. 510-516, 2005.

MATTHEWS, B. W. Solvent content of protein crystals. **Journal of Molecular Biology**, v. 33, p. 491-497, 1968.

MURSHUDOV, G. N.; SKUBÁK, P.; LEBEDEV, A. A.; PANNU, N. S.; STEINER, R. A.; NICHOLLS, R. A.; WINN, M. D.; LONG, F.; VAGIN, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. **Acta Crystallographica Section D**, p. 355-367, 2011.

NAM, K. H.; KIM, M-Y.; KIM, S-J.; PRIYADARSHI, A.; KWON, S-T.; KOO, B-S.; YOON, S-H.; KWANG, K. Y. Structural and functional analysis of a novel hormone-sensitive lipase from a metagenome library. **Proteins**, v. 74, p. 1036-1040, 2009.

NARDINI, M.; DIJKSTRA, B. W.  $\alpha/\beta$  Hydrolase fold enzymes: the family keeps growing. **Current opinion in Structural Biology**, v. 9, p. 732-737, 1999.

NGO, T. D.; RYU, B. H.; JU, H.; JANG, E.; PARK, K.; KIM, K. K.; KIM, T. D. Structural and functional analyses of a bacterial homologue of hormone-sensitive lipase from a metagenome library. **Acta Crystallographica Section D, Biological Crystallography**, p. 1726-1737, 2013.

OLLIS, D. L.; CHEAH, E.; CYGLER, M.; DIJKSTRA, B.; FROLOW, F.; FRANKEN, S. M.; HAREL, M.; REMINGTON, S. J.; SILMAN, I.; SCHRAG, J. The alpha/beta hydrolase fold. **Protein Engineering, Design and Selection**, v. 5, n. 3, p. 197-211, 1992.

PENG, Q.; WANG, X.; SHANG, M.; HUANG, J.; GUAN, G.; LI, Y.; SHI, B. Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. **Microbial Cell Factories**, v. 13, n. 1, 2014.

PERÄNEN J., RIKKONEN M., HYVÖNEN M., KÄÄRIÄINEN L. T7 vectors with modified T7 lac promoter for expression of proteins in *Escherichia coli*. **Analytical Biochemistry**, p. 371-373, 1996.

QUAX, W.J.; BROEKHUIZEN, C.P. Development of a new Bacillus carboxyl esterase for use in the resolution of chiral drugs. **Applied Microbiology and Biotechnology**, v. 41, n. 4, p. 425-431, 1994.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain conformations. **Journal of Molecular Biology**, v. 7, p. 95-99, 1963.

RAO, L.; XUE, Y.; ZHENG, Y.; LU, J. R.; MA, Y. A novel alkaliphilic Bacillus esterase belongs to the 13<sup>th</sup> bacterial lipolytic enzyme family. **PLoS One**, v. 8, 2013.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, p. 406-425, 1987.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Molecular Biology and Evolution**, v. 24, p. 1596-1599, 2007.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, p. 4673-4680, 1994.

WEI, Y.; CONTRERAS, J. A.; OSTERLUND, T.; DEREWENDA, U.; KNEUSEL, R. E.; MATERN, U.; DEREWENDA, Z. S. Crystal structure of brefeldin A esterase, a bacterial homolog of the mammalian hormone sensitive lipase. **Nature Structural Biology**, v. 6, n. 4, p. 340-345, 1999.

WINN, M. D.; BALLARD, C. C.; COWTAN, K. D.; DODSON, E. J.; EMSLEY, P.; EVANS, P. R.; KEEGAN, R. M.; KRISINEL, E. B.; LESLIE, A. G. W.; MCCOY, A.; MCNICHOLAS, S. J.; MURSHUDOV, G. N.; PANNU, N. S.; POTTERTON, E. A.; POWELL, H. R.; READ, R. J.; VAGIN, A.; WILSON, K. S. Overview of the CCP4 suite and

current developments. *Acta Crystallographica Section D, Biological Crystallography*, p. 235-242, 2011.

#### 4.3.8 Supplementary material

**Table S1 - Proteins selected to build the phylogenetic tree.** Two members from each family of bacterial lipolytic enzymes were selected and the sequences extracted from NCBI (<http://www.ncbi.nlm.nih.gov>). Also, comparing these sequences against the Protein Data Bank (PDB) through blastp, it was possible to identify proteins with structure solved; wherein, for each family a PDB code was selected. The phylogenetic tree was built using MEGA 4.1 (TAMURA et al., 2007) after amino acid sequence alignment using ClustalW (Thompson et al., 1994).

Genbank or PDB code	Microorganism	Protein function	Blast X PDB Query cover/identity (%)
AAA50466.1	<i>Burkholderia cepacia</i>	Lipase	1OIL 87/94
AAC05510.1	<i>Pseudomonas luteola</i>	Triacylglycerol lipase precursor	1OIL 88/93
1OIL	<i>Pseudomonas cepacia</i>	Triacylglycerol lipase	1OIL 100/100
AAC38796.1	<i>Salmonella typhimurium</i>	Outer membrane esterase	3KVN 75/28
CAA47020.1	<i>Photobacterium luminescens</i>	Triacylglycerol lipase	3KVN 80/26
3KVN	<i>Pseudomonas aeruginosa</i>	Full-length autotransporter EstA	3KVN 100/100
AAA53485.1	<i>Streptomyces albus</i>	Lipase precursor	1JFR 85/83
AAB51445.1	<i>Streptomyces</i> sp.	Triacylglycerol acylhydrolase	1JFR 84/100
1JFR	<i>Streptomyces exfoliatus</i>	Lipase	1JFR 100/100
AAC41424.1	<i>Cupriavidus necator</i>	Lipase-like enzyme	1QZ3 86/37
CAA37862.1	<i>Moraxella</i> sp.	Triacylglycerol lipase	1JJI 76/32

2YH2	<i>Pyrobaculum calidifontis</i>	Esterase	2YH2 100/100
AAA25933.1	<i>Pseudomonas oleovorans</i>	PHA-depolymerase	2VF2 78/27
CAA47949.1	<i>Psychrobacter immobilis</i>	Triacylglycerol lipase	4OPM 87/34
1Q0R	<i>Streptomyces purpurascens</i>	Aclacinomycin methylesterase	1Q0R 100/100
AAC67727.1	<i>Chlamydia trachomatis</i>	Lysophospholipase esterase	4F21 84/24
AAB30793.1	<i>Arthrospira platensi</i>	Serine esterase	4FHZ 97/32
1AUO	<i>Pseudomonas fluorescens</i>	Carboxylesterase	1AUO 100/100
Q01470.1	<i>Arthrobacter oxydans</i>	Phenmedipham hydrolase	2OGT 95/36
P37967.2	<i>Bacillus subtilis</i>	Para-nitrobenzyl esterase	1QE3 100/98
1QE3	<i>Bacillus subtilis</i>	Esterase	1QE3 100/100
CAA78842.1	<i>Streptomyces anulatus</i>	Esterase A	3ZYT 89/34
AAA99492.1	<i>Arthrobacter globiformis</i>	Carboxylic ester hydrolase	3ZYT 99/81
3ZYT	<i>Arthrobacter</i> sp.	Esterase	3ZYT 100/100

**Table S2 - Data collection and refinement statistics of Est8.**

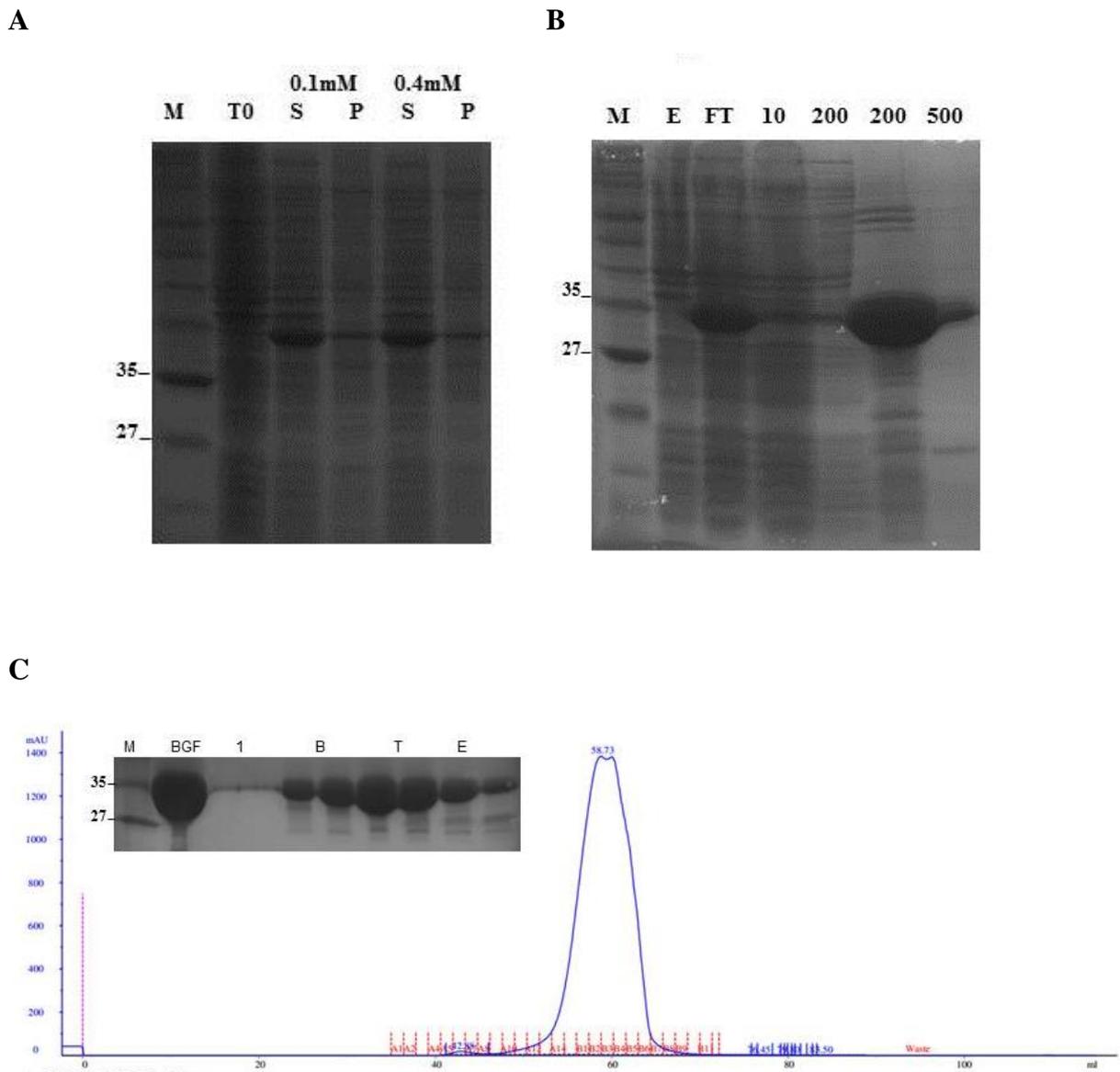
Contents	Est8
<i>Data collection</i>	
Beamline	Diamond I02
Wavelength	0.919150 Å
Resolution	30.000 to 1.850 Å
Space group	P 3 <sub>1</sub> 2 1
Cell dimensions (Å) <i>a, b, c</i>	85.50 85.50 98.67 90.000 90.000 120.000
R <sub>merge</sub> (%)	4.6 (57.1)

Mean $I/\sigma(I)$	14.44 (2.23)
$R_{\text{meas}}$ (%)	5.4 (67.2)
CC(1/2)	99.9% (81.9%)
Redundancy	3.53
<i>Refinement statistics</i>	
Resolution	27.985 - 1.850
No. of reflections	35626
$R_{\text{work}}/R_{\text{free}}$ (%)	17.1/ 18.91
No. of protein atoms	2531
<i>R.m.s deviations</i>	
Bond lengths (Å)	0.006
Bond angles (°)	1.024
<i>Ramachandran distribution</i>	
Favoured (%)	95.5
Allowed (%)	3.22
Disallowed (%)	1.29

#### 4.4 Chapter 4: Est30 protein

##### 4.4.1 Expression and purification

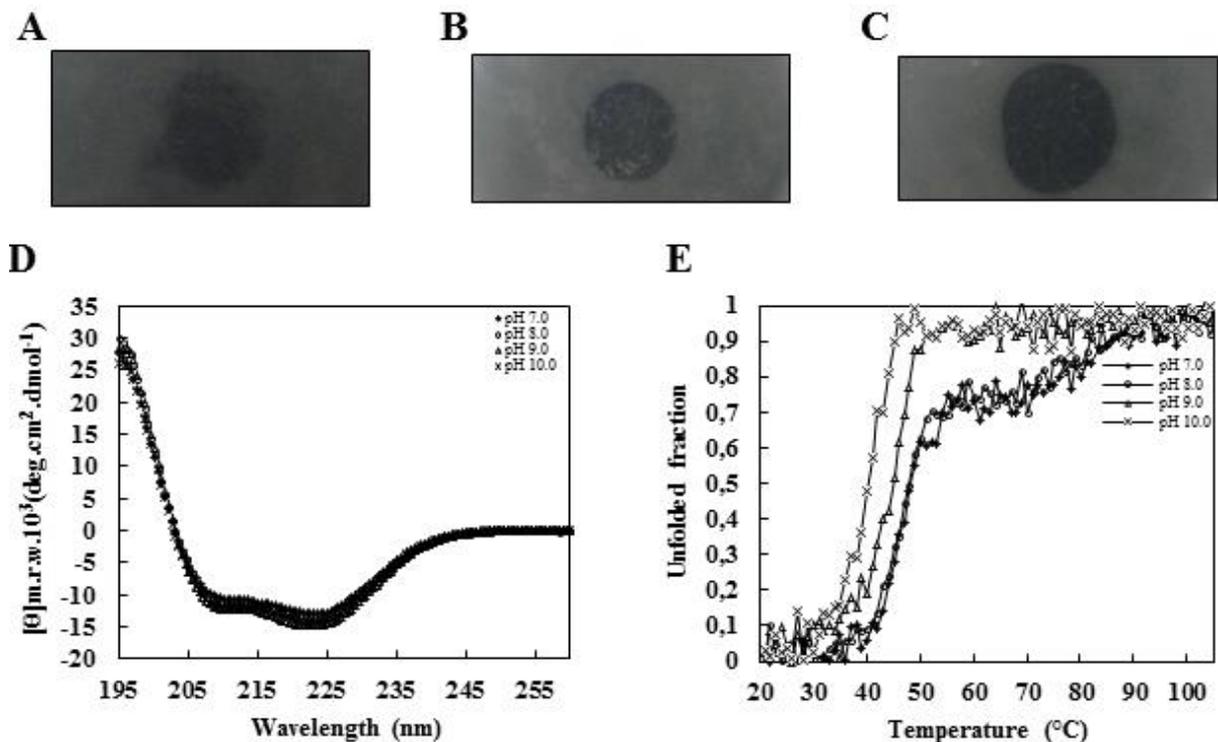
The *est30* gene was cloned into the pHAT2 vector to create the pHAT2-*est30* construct. This construct was submitted to expression tests and the temperature and IPTG concentrations were varied (Figure 7A), but no differences were observed between the two concentrations of IPTG. The Est30 protein was expressed in a soluble form with a molecular weight of 35 kDa, which was expected from the amino acid sequence. The purification was carried out using immobilized metal (nickel) affinity chromatography and the protein was eluted with buffer containing 200 mM imidazole. The resulting eluent still contained some contaminants (Figure 7B). To eliminate the contaminants, the purified samples were subjected to molecular size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column (Figure 7C). The results showed that the Est30 protein was eluted at a column volume of 58.74 mL, with a reduction in the amount of contaminants (Figure 7C, inset).



**Figure 7. Expression and purification of the Est30 protein.** (A) Expression tests of Est30. The *E. coli* BL21 (DE3) cells carrying the pHAT-*est30* construct were induced with 0.1 mM or 0.4 mM IPTG for 20 hours at 20 °C. M: protein marker (New England BioLabs); T0: non-induced cells; S: soluble fraction; P: insoluble fraction. (B) The results of the immobilized metal ion affinity chromatography. M: protein marker (New England BioLabs); E: soluble extracts; FT: flow through; 10, 200, and 500: wash samples with buffer containing 10 mM, 200 mM, and 500 mM imidazole, respectively. (C) Est30 chromatogram obtained after molecular size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare Bio-Sciences). The inset polyacrylamide gel on the left shows the following: (M) protein marker (New England BioLabs); (BGF) sample from before gel filtration; (1) samples from small peak; samples from the beginning (B), the center (T) and the end (E) of the large peak.

#### 4.4.2 Assay in Petri dishes containing tributyrin and circular dichroism

The activity of purified Est30 was preliminarily evaluated against tributyrin using Petri dishes assays with different pH values. Based on the halo formation, the results showed that Est30 cleaves tributyrin more efficiently in basic pH (8.0 and 9.0) than in neutral pH (Figure 8A-C). On the other hand, the loss of activity was not related to differences in the secondary structure of Est30, which revealed an alpha-beta structure in the circular dichroism spectra obtained for the samples incubated at different pH from 7.0 to 10.0 (Figure 8D). In addition, the thermal denaturation experiments carried out at a constant pH revealed that Est30 is a mesophilic enzyme with a melting temperature of approximately 42-47 °C (Figure 8E).



**Figure 8. Tributyrin assay performed in Petri dishes and circular dichroism of Est30.** A tributyrin assay of Est30 (50  $\mu\text{g}/\text{mL}$ ) performed in Petri dishes containing 0.23% Tris-HCl (w/v); agar 1.2% (w/v) and 1% tributyrin is shown at pH 7.0 (A), pH 8.0 (B) and pH 9.0 (C). (D) Circular dichroism spectra of Est30 at pH 7.0 ( $\blacklozenge$ ), pH 8.0 ( $\circ$ ), pH 9.0 ( $\Delta$ ) and pH 10.0 ( $\times$ ). Samples containing 0.2 mg/mL of protein (corresponding to 5.5  $\mu\text{mol}$ ) were prepared and the spectra were recorded using a Jasco J-810 spectropolarimeter. Similarly, the thermal denaturation of the enzyme was measured at 222 nm as a function of the temperature at pH 7.0 ( $\blacklozenge$ ), pH 8.0 ( $\circ$ ), pH 9.0 ( $\Delta$ ) and pH 10.0 ( $\times$ ).

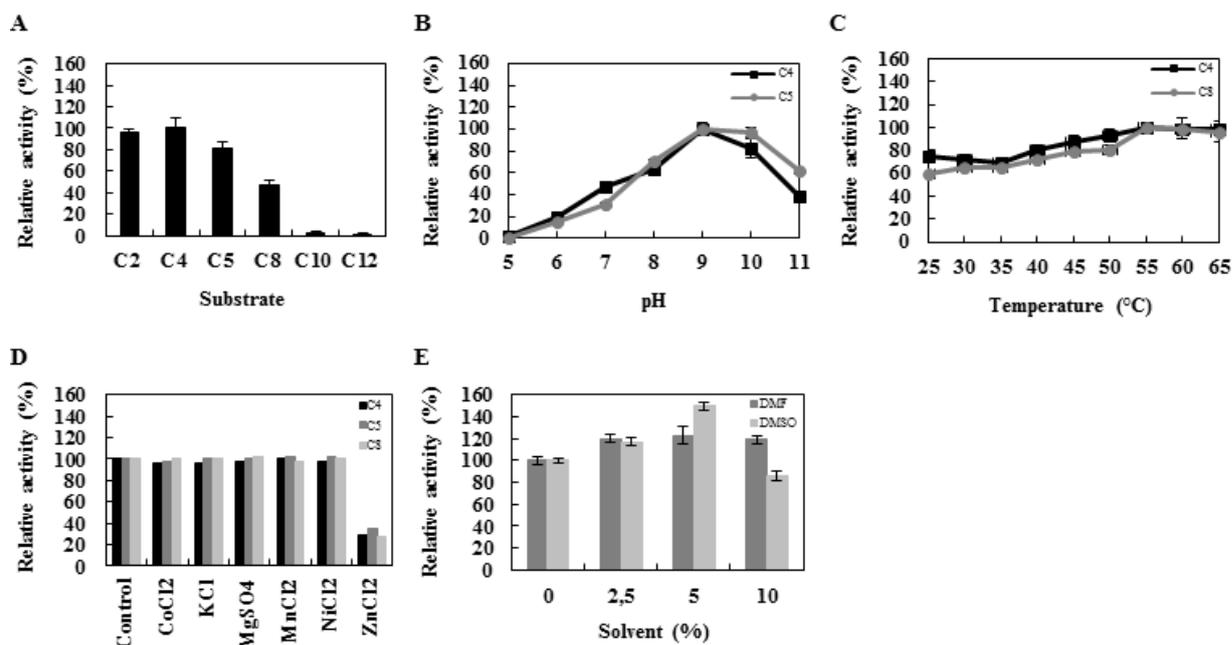
#### 4.4.3 Substrate specificity and kinetic parameters of Est30

The substrate specificity of Est30 was assessed using *p*-nitrophenyl (*p*-NP) esters with different acyl chain lengths (C<sub>2</sub> to C<sub>10</sub>). Est30 was able to hydrolyze substrates with acyl chains up to 12 carbons (Figure 9A) and had high catalytic efficiency (K<sub>cat</sub>.K<sub>m</sub><sup>-1</sup>) toward *p*-NP valerate (C<sub>5</sub>) (Table 4). A sharp decrease in the activity was observed for acyl chain lengths with more than 10 carbons. *p*-NP butyrate and valerate were substrates with higher catalytic efficiencies of approximately 74.2 and 113.9 min<sup>-1</sup>.μM<sup>-1</sup>, respectively (Table 4). The catalytic efficiency toward butyrate was 2.2 and 12.4 times higher than the catalytic efficiency toward octanoate and decanoate, respectively, whereas the catalytic efficiency toward valerate was 1.5, 3.4 and 19.6 times higher than the catalytic efficiency toward butyrate, octanoate and decanoate, respectively. The preference of Est30 for short chain substrates (C < 10) justified its designation as an esterase.

Using 1 mM of *p*-NP butyrate (C<sub>4</sub>) or *p*-NP octanoate (C<sub>8</sub>) as substrates, the Est30 was shown to be an alkaline esterase with a maximum activity at pH 9.0 (Figure 9B). At pH 10.0, Est30 still presented high activity when compared to pH 9.0, showing approximately 82% and 97% of relative activity against *p*-NP butyrate (C<sub>4</sub>) and *p*-NP octanoate (C<sub>8</sub>), respectively. Below pH 9.0, Est30 displayed a gradual decrease in the activity. This behavior is in agreement with the results of the tributyrin assay in Petri dishes, which indicated the highest activity of Est30 at pH 9.0 (Figure 8C).

The temperature effect on the hydrolysis of *p*-NP butyrate and *p*-NP octanoate by Est30 was observed at temperatures below 50 °C (Figure 9C). In fact, the protein showed high activity at temperatures of 60 °C to 65 °C for both substrates and could be applied in biotechnology as a thermal enzyme.

The effects of metal ions on the enzymatic activity were evaluated by measuring the residual esterase activity in the presence of 0.5 mM of CoCl<sub>2</sub>, KCl, CaCl<sub>2</sub>, MgSO<sub>4</sub>, MnCl<sub>2</sub> or ZnCl<sub>2</sub>. The enzyme activity did not show any change in the presence of metals except for in the presence of Zn<sup>+2</sup>, which induced a 65% reduction in the activity of the enzyme in comparison to the control (Figure 9D). To test the effect of solvents on the activity of Est30, assays were performed in the presence of DMF and DMSO. The activity of Est30 increased in the presence of DMF and in the presence of up to 5% DMSO when compared to the control.



**Figure 9. Kinetics analyses of Est30.** (A) Substrate specificity against *p*-nitrophenyl esters with different acyl chain lengths (C<sub>2</sub> to C<sub>12</sub>); (B) Effect of pH on Est30 activity against *p*-nitrophenyl butyrate (■) and *p*-nitrophenyl valerate (◆); (C) Effect of temperature on Est30 activity against *p*-nitrophenyl butyrate (■) and *p*-nitrophenyl valerate (◆); (D) Activity of Est30 against *p*-nitrophenyl butyrate (■), *p*-nitrophenyl valerate (■) and *p*-nitrophenyl octanoate (■) in the presence of different cations; (E) The influence of organic solvents, DMF (■) and DMSO (■) on the activity of Est30 against *p*-nitrophenyl butyrate.

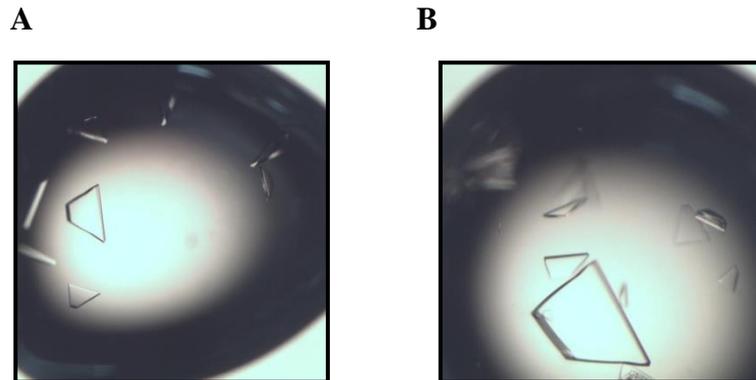
**Table 4 - Kinetic parameters of Est30 activity against *p*-nitrophenyl substrates.**

Substrate	K <sub>m</sub> ( $\mu$ M)	V <sub>max</sub> ( $\mu$ M.min <sup>-1</sup> )	K <sub>cat</sub> (min <sup>-1</sup> )	K <sub>cat</sub> .K <sub>m</sub> <sup>-1</sup> (min <sup>-1</sup> . $\mu$ M <sup>-1</sup> )
<i>p</i> -nitrophenyl acetate	229.6 $\pm$ 23.9	76.07 $\pm$ 2.5	4876.2 $\pm$ 160.2	21.2 $\pm$ 6.7
<i>p</i> -nitrophenyl butyrate	172.7 $\pm$ 32.9	200.0 $\pm$ 11.2	12820.5 $\pm$ 717.9	74.2 $\pm$ 21.8
<i>p</i> -nitrophenyl valerate	100.1 $\pm$ 9.2	177.9 $\pm$ 4.3	11403.8 $\pm$ 275.6	113.9 $\pm$ 29.9
<i>p</i> -nitrophenyl octanoate	82.7 $\pm$ 30.8	42.7 $\pm$ 3.9	2737.1 $\pm$ 250.0	33.1 $\pm$ 8.1
<i>p</i> -nitrophenyl decanoate	100.5 $\pm$ 4.3	9.2 $\pm$ 0.1	589.7 $\pm$ 6.4	5.8 $\pm$ 1.5

#### 4.4.4 Crystallization trials and diffraction

The pure Est30 was used in seven crystallization screens, including Classic Suite (Qiagen), JCSG Suite (Qiagen), Midas (Molecular Dimensions), Morpheus (Molecular Dimensions), PEGsI (Qiagen), pHClear I (Qiagen), and Wizard I & II (Emerald BioSystems). In total, 672 conditions were screened with 15.3 mg/mL of the protein. The following two conditions resulted in crystal formation: 1.6 M ammonium sulfate, 0.1 M HEPES pH 7.5, 0.1 M sodium chloride from the Classic Suite crystallization kit (well C08); and 1.6 M

ammonium sulfate, 0.1 M bicine pH 9.0 from pH Clear crystallization kit (well E12). These crystals were refined by varying the pH and concentration of the precipitation agent. After refinement, diffraction-quality crystals were obtained with 1.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, and 0.1 M sodium chloride, or with 1.6 M ammonium sulfate, 0.1 M HEPES pH 8.2, and 0.1 M sodium chloride (Figure 10).



**Figure 10. Est30 Crystals.** The crystals of Est30 protein were obtained in 1.2 M ammonium sulfate, 0.1 M HEPES pH 7.5, 0.1 M sodium chloride (A) and 1.6 M ammonium sulfate, 0.1 M HEPES pH 8.2, 0.1 M sodium chloride (B). The protein was crystallized at 15.3 mg/ml using the hanging-drop method (1  $\mu$ l protein/0.5  $\mu$ l reservoir solution).

The 1.70 Å resolution crystal structure of Est30 was determined in the centered monoclinic space group C2, with three molecules present in the asymmetric unit. Several cycles of maximum likelihood refinement using phenix.refine and REFMAC, followed by manual model rebuilding, reduced the  $R$  values for all data. The  $R$  factor of the present model was 17.21%, with an  $R_{\text{free}}$  of 20.33%. The crystallographic statistics for data collection and structure refinement are summarized in Table 5.

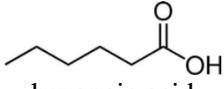
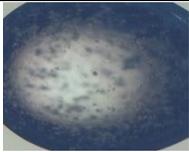
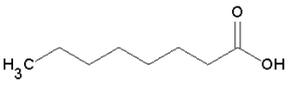
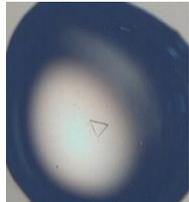
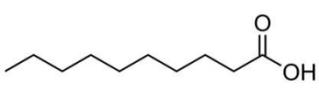
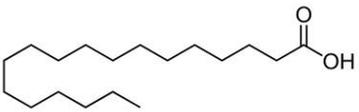
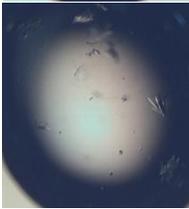
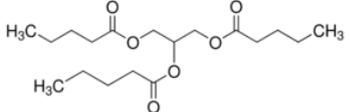
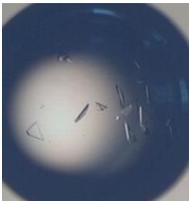
**Table 5 – Data collection and refinement statistics for Est30.**

		<b>Est30</b>				
Space group		C2 (5)				
Cell dimensions (Å) $a, b, c$		182.06	45.67	135.27	90.000	105.449 90.000
Resolution (Å)		130.05 - 1.70				
$R_{\text{merge}}$ (%)		18.3 (53.0)				
$I/\sigma I$		7.42 (1.79)				
$R_{\text{meas}}$		21.7 (62.8)				
CC(1/2)		96.6 (74.2)				
Redundancy		0.29				
<b>Refinement</b>						
Resolution		130.05 - 1.7				
No. of reflections		110117				
$R_{\text{work}}/R_{\text{free}}$ (%)		17.21/ 20.33				
Total no. of atoms		7273				

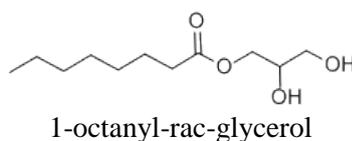
<b>r.m.s. deviations</b>	
Bond lengths (Å)	0.006
Bond angles (°)	1.011
<b>Ramachandran plot</b>	
Favored (%)	95.24
Allowed (%)	3.81
Disallowed (%)	0.95

In addition, the Est30 crystals were soaked with sodium butyrate and co-crystallized with different substrates, including hexanoic acid, octanoic acid, decanoic acid, stearic acid, glycerol trivalerate and 1-octanyl-rac-glycerol (Table 6). The diffraction data generated using sodium butyrate was processed and the structure was determined in the same space group (C2 at 1.65 Å). The statistics of the data collection and refinement of the Est30-butyrate complex are presented in Table 7.

**Table 6 - Crystallization conditions of Est30 protein and ligands used for co-crystallization.**

Protein concentration (mg/mL)	Crystallization condition	Ligand	Est30 crystal in visible light
10	1.2 M ammonium sulfate 0.1 M bicine pH 8.0	 hexanoic acid	
10	1.2 M ammonium sulfate 0.1 M bicine pH 8.0	 octanoic acid	
10	1.4 M ammonium sulfate 0.1 M bicine pH 8.0	 decanoic acid	
10	1.2 M ammonium sulfate 0.1 M bicine pH 8.0	 stearic acid	
15	1.2 M ammonium sulfate 0.1 M bicine pH 8.0	 glycerol trivalerate	

15

1.2 M ammonium Sulfate  
0.1 M bicine pH 8.0**Table 7 – Data collection and refinement statistics for the Est30-butyrate complex.**

Data collection	Est30 + butyrate					
Space group	C2 (5)					
Cell dimensions (Å) <i>a, b, c</i>	181.82	45.42	134.22	90.000	105.325	90.000
Wavelength	0.919150					
Resolution	129.45 - 1.65					
R <sub>merge</sub> (%)	8.1 (106.3)					
<i>I</i> / $\sigma$ <i>I</i>	11.49 (2.01)					
R <sub>meas</sub> (%)	9.1 (124.1)					
CC(1/2)	99.5% (49.2%)					
Redundancy	4.88					
<b>Refinement</b>						
Resolution (Å)	129.45 - 1.65					
No. of reflections	118723					
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	22.05/24.53					
Total no. of atoms	7385					
<b>r.m.s. deviations</b>						
Bond lengths (Å)	0.022					
Bond angles (°)	2.325					
<b>Ramachandran plot</b>						
Favored (%)	93.25					
Allowed (%)	4.39					
Disallowed (%)	2.36					

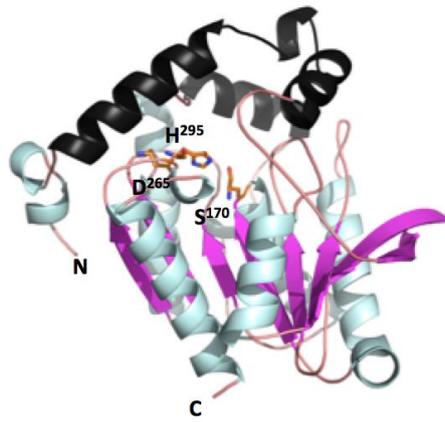
Est30 has a characteristic  $\alpha/\beta$ -fold consisting of one  $\beta$ -sheet composed of 8  $\beta$ -strands (only strand  $\beta$ 2 is antiparallel) surrounded by twelve helices (Pro9-Phe12; Pro14-Ala16; Asp19-32Ala; Trp38-Ile41; Ala43-Arg52; Asp110-Ala120; Ala139-Phe159; Ala171-His187; Thr222-Phe232; Pro246-Tyr248; Val268-Ala281; Glu303-Thr322). The Est30 protein also has a cap domain composed of the residues Asp19-Arg52 and Thr222-Phe231 (Figure 11A). The highly conserved catalytic triad of Est30 is defined by the residues Ser170 (located between  $\beta$ 5 and  $\alpha$ 8), Asp265 (located between  $\beta$ 7 and  $\alpha$ 11) and His295 (located between  $\beta$ 8 and  $\alpha$ 12; see Figure 11A, orange sticks). Though the three amino acids in the catalytic triad were located in close proximity, Asp265 and His295 were especially close to each other (Figure 11A).

The structure of Est30 shows a tunnel through the protein, which may possibly allow the substrate to enter into the active site through one side followed by delivery of the product

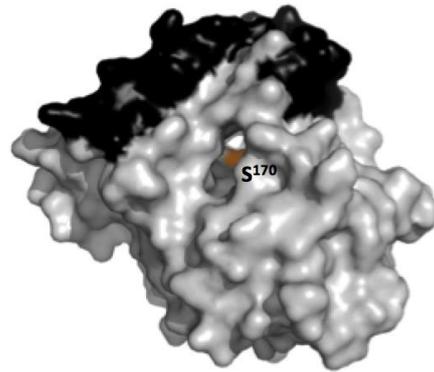
through the opposite side (Figure 11B-C). The structure of the Est30-butyrate complex revealed that the substrate was located inside the tunnel and close to the catalytic residues Asp265 and His295 (Figure 11D). In addition, it was possible to identify the formation of an oxyanion hole mediated by Gly98, Gly99 and Gly100 that was located in the <sup>97</sup>HGGG<sup>100</sup> sequence motif, which is located 73 residues upstream of Ser170 in the active site. The two other motifs that describe members of family IV, <sup>167</sup>GGESAGGHL<sup>175</sup> and <sup>265</sup>DALVDD<sup>270</sup>, are also shown in Figure 11E.

Structural superposition between free Est30 and the Est30-butyrate complex showed no differences in the cap conformation (Figure 11F). Rengachari and collaborators (2013) have shown that the cap domain in the monoacylglycerol lipase (bMGL) presented conformational plasticity when the protein was in the free form (super-open conformation) or in a complex (conformation II or III). It is possible that this cap domain plasticity was not observed in Est30 due to the short chain length of the substrate relative to the substrates shown in the bMGL structures. However, this feature would be interesting to evaluate using others Est30 structures in complex with substrates that have long chain lengths.

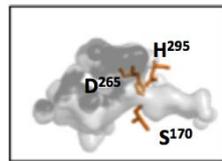
A



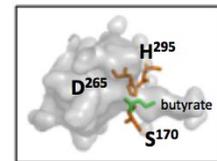
B



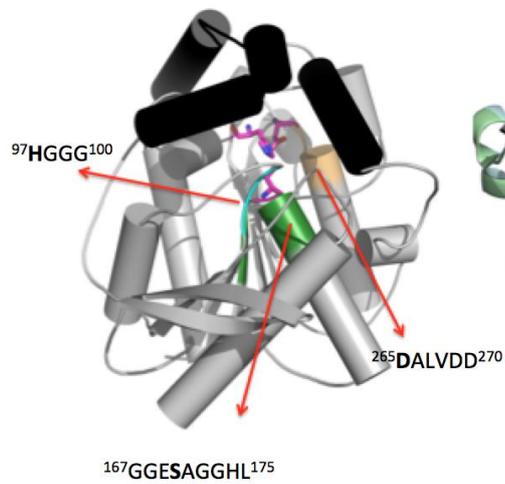
C



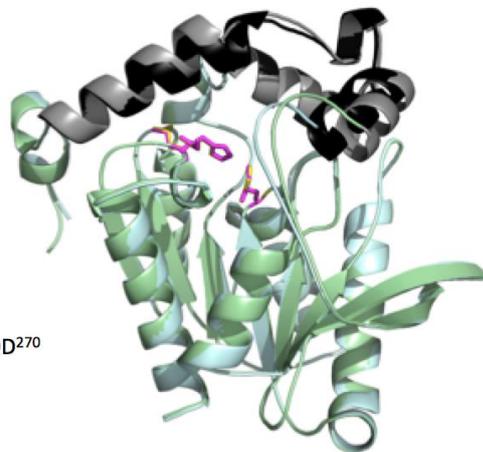
D



E



F



**Figure 11. Three-dimensional structure of Est30.** (A) A cartoon representation of the structure with helices in pale cyan, beta-sheets in magenta and loops in sand. Residues forming the catalytic triad are colored as orange sticks, and the helices forming the domain cap are colored in black. (B) A surface view of Est30 in gray showing the detached opening channel for the substrate and the passage of the product. Ser170 from the catalytic triad is shown in orange inside the channel. (C) The full channel is shown for Est30 and the red arrow shows the direction of substrate's entrance. (D) Est30, with butyrate shown as a green stick. (C-D) The insets in the figure show the volume of the pocket in the surface. (E) The structure of the Est30 evidencing the three motifs from family IV that support the residues of the catalytic triad. (F) Structural superposition of free Est30 (pale blue) and butyrate-bound Est30 (pale green), evidencing the small differences in the cap domains (gray in Est30 and black in Est30 + butyrate). The residues from the catalytic triad did not show conformational changes and are evidenced as pink sticks.

#### 4.4.5 Construction of *Est30asp19* and *Est30ala34* mutants

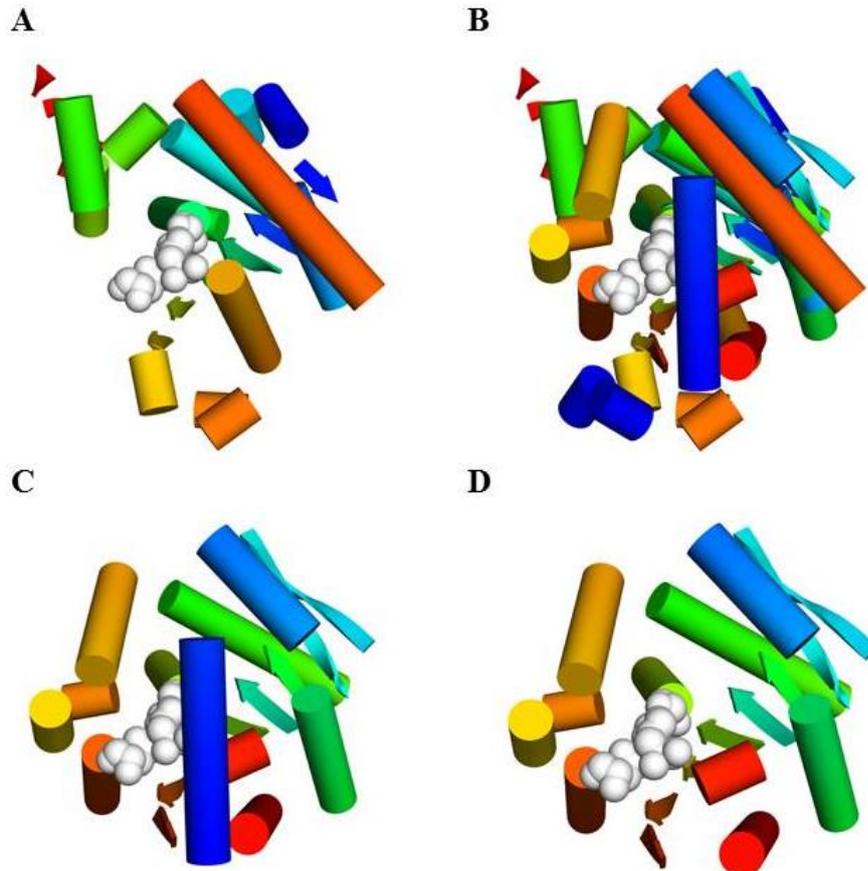
As demonstrated in figure 6, the Est30 protein was located in the same branch as a lipase from *Bacillus* sp. used in biodiesel production (patent number WO2008032007).

By comparing the structure of Est30 to the solved structures of other lipases, it was possible to identify lipase B from *Candida antarctica*, an enzyme with a exposed active site that displays a stereospecificity pocket for secondary alcohols (Figure 12A) (UPPENBERG et al., 1995). The sequence alignment between Est30 and lipase B showed a 9.7% amino acid sequence identity and the superposition of both structures showed quite similar folding patterns, with the exception of three additional  $\alpha$ -helices in Est30 enclosing the active site (Figure 12B, in dark blue).

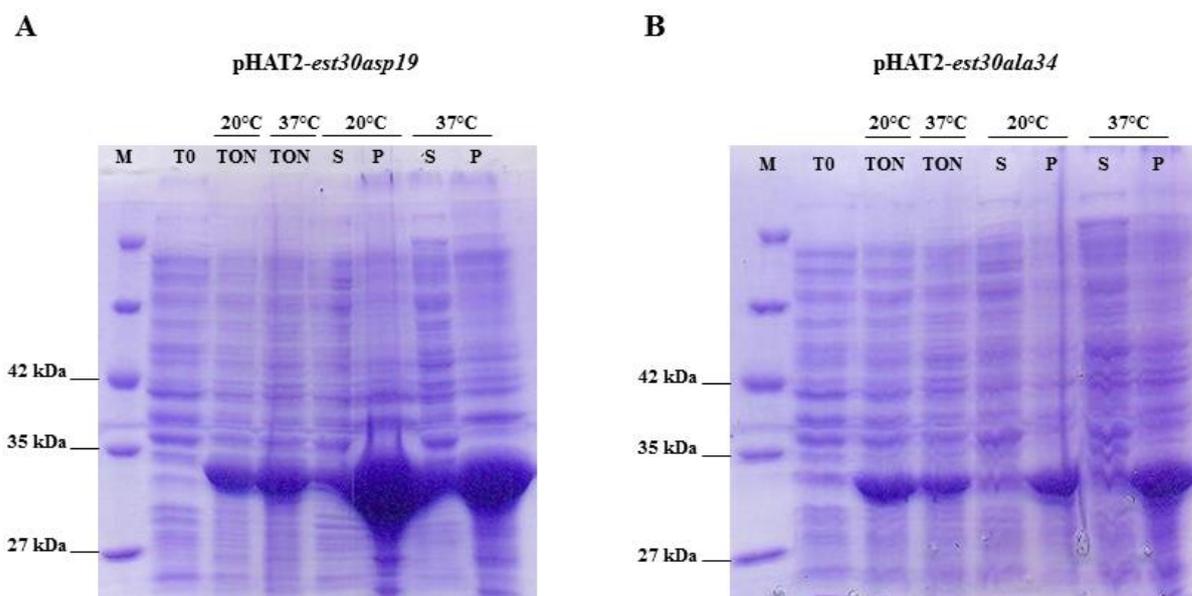
In view of this result, two distinct mutants of Est30 were designed with more open active sites. The first two helices (amino acids 1 to 18 in the N-terminal) were deleted from the first mutant, whereas the first three helices (amino acids 1 to 33 in the N-terminal) were deleted from the second mutant (Figure 12C and D, respectively). The goal was to verify the importance of these helices for the enzyme activity, and it was expected that exposing the substrate-binding site would decrease the energy of activation required to carry out the reactions. Indeed, it should be noted that these mutations would enable easy access of a wide range of substrates directly to the residues of the catalytic triad.

The two mutant recombinant proteins were designated Est30asp19 and Est30ala34, and the *est30asp19* and *est30ala34* genes were cloned individually into the pHAT2 vector in order to create the pHAT2-*est30asp19* and pHAT2-*est30ala34* constructs, respectively. These constructs were submitted to expression tests in which the temperature was varied, being possible to verify that the corresponding proteins were expressed with molecular weights of

34 and 32.3 kDa, respectively (which were expected from the amino acid sequence analyses performed using the ProtParam tool). However, both proteins were expressed as insoluble forms (Figure 13) and the yield of the soluble fraction was practically zero. It is likely that the amino acids removed in each construct affected the protein solubility. Additional protocols and methods will be tested in order to make these constructs viable.



**Figure 12. Structural comparison of Est30 and Lipase B from *Candida antarctica* (PDB code 1LBS).** Helices are shown in cylinders, the catalytic triad is shown in white dots and the helices are depicted in blue. (A) The structure of Lipase B and its exposed active site. (B) The structural superposition of Est30 and Lipase B illustrating the three additional helices in the N-terminal of Est30 (dark blue cylinders). Representations of the three-dimensional structure of the Est30 protein with 18 (C) or 33 (D) residues deleted from the N-terminus. The mutants were designated Est30asp19 and Est30ala34, respectively.

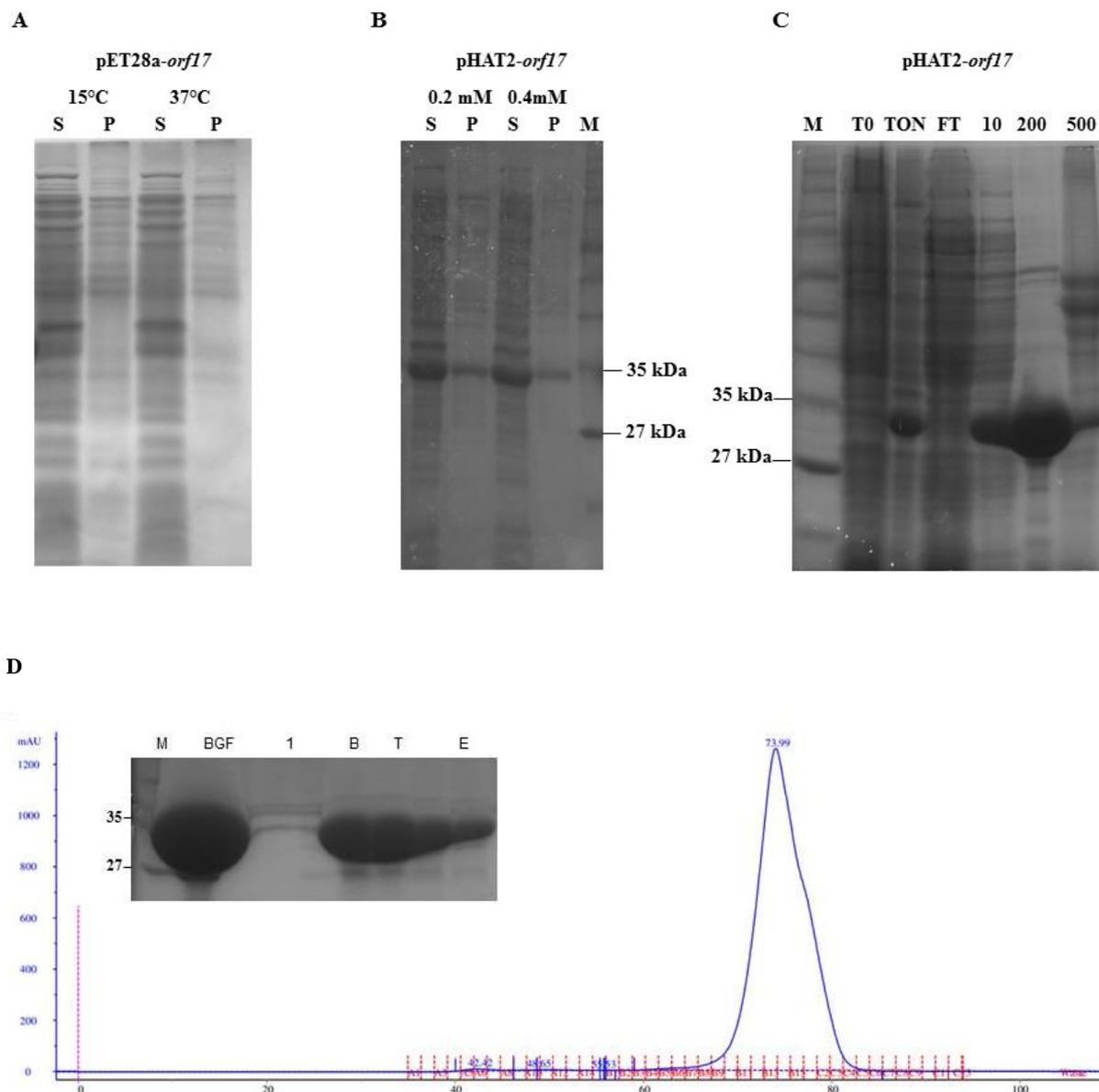


**Figure 13. Expression test of pHAT2-est30asp19 and pHAT2-est30ala34 constructs.** The *E. coli* BL21 (DE3) cells carrying the pHAT-est30asp19 (A) or pHAT-est30ala34 (B) constructs were induced with 0.4 mM IPTG for 20 hours at 20 °C or for 4 hours at 37 °C. M: protein marker (New England Biolabs); T0: non-induced cells; TON: induced cells; S: soluble fraction; P: insoluble fraction.

## 4.5 Chapter 5: ORF17 protein

### 4.5.1 Expression and purification

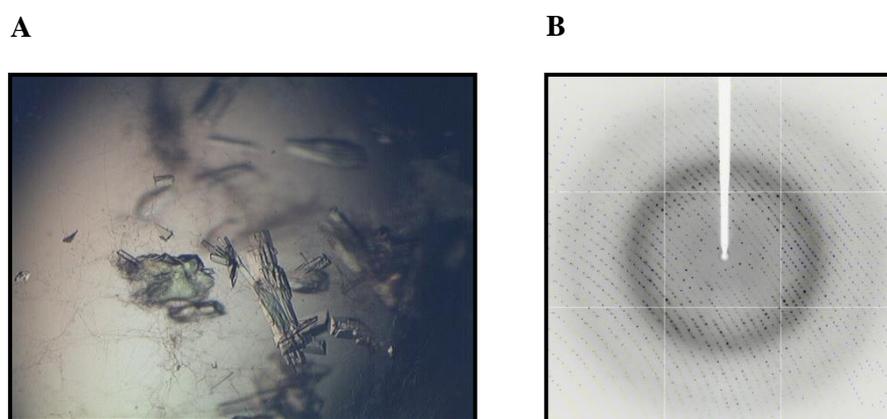
*ORF17* was poorly expressed as a soluble protein from *E. coli* cells carrying the pET28a-*orf17* vector at either 15 °C or 37 °C (Figure 14A), though the expression of soluble protein from the pHAT2-*orf17* vector increased by two-fold (Figure 14B). No significant difference was noticed when 0.2 mM or 0.4 mM IPTG was used. The large-scale expression of the *ORF17* protein was performed at 15 °C for 20 hours with 0.4 mM IPTG. After the protein extraction, most of the *ORF17* protein was eluted using a buffer containing 200 mM (Figure 14C). The size exclusion chromatography of the *ORF17* protein showed a single peak that was eluted at a column volume of 74 mL (Figure 14D). Then, the samples were pooled, concentrated, and used for crystallization trials.



**Figure 14. Expression and purification of the *ORF17* protein.** The *E. coli* BL21 (DE3) cells carrying the pET28a (A) or pHAT2 (B) vectors containing the *orf17* gene were submitted to expression tests. In (A), the pET28a-*orf17* construct was tested at 15 °C and 37 °C using 0.4 mM IPTG; in (B), the pHAT2-*orf17* construct was tested at 37 °C using 0.1 mM or 0.4 mM IPTG. M: protein marker (New England BioLabs); S: soluble fraction; P: insoluble fraction. (C) The polyacrylamide gel showing the *ORF17* purification with IMAC. M: protein marker (New England BioLabs); T0: non-induced cells; TON: cells induced at 15 °C using 0.4 mM IPTG; E: soluble extracts; FT: flow through; 10, 200 and 500: wash samples with buffer containing 10 mM, 200 mM and 500 mM imidazole, respectively; (D) Chromatogram profile obtained after molecular size-exclusion chromatography of *ORF17* using a HiLoad 16/60 Superdex 75 column (GE Healthcare Bio-Sciences). The inset polyacrylamide gel on the left shows the following: (M) protein marker (New England BioLabs); (BGF) sample before gel filtration; samples from the beginning (B), the center (T) and the end (E) of the large peak.

### 4.5.2 Crystallization

The crystallization screen of *ORF17* was performed at the Department of Biochemistry at the University of Cambridge. A total of 384 different crystallization conditions were tested with a final protein concentration of 14 mg/mL, 32 mg/mL and 65 mg/mL. The best condition for crystal growth was 0.1 M HEPES at pH 7.0 and 30% PEG 6000 (from pH Clear I, well D10). This condition was reproduced in large drops and refined by varying the pH and the concentration of the precipitating agent. From the crystals subjected to the beamline, only one that was grown in 0.1 M HEPES at pH 7.0 and 28% PEG 6000 and with 65 mg/mL of pure protein generated a good diffraction pattern (Figure 15).



**Figure 15. Crystallization and diffraction of the *ORF17* protein.** (A) *ORF17* crystals were obtained in 0.1 M HEPES at pH 7.0 and 28% PEG 6000. The protein was crystallized at 65 mg/ml using the hanging-drop method (1  $\mu$ L protein/1  $\mu$ L reservoir solution). (B) Diffraction pattern of *ORF17* crystals collected on the beamline ID14 at the European Synchrotron Radiation Facility (ESRF) using a Q315r ADSC X-ray detector.

The 1.4 Å resolution crystal structure of *ORF17* was determined in space group  $P2_12_12_1$  with two molecules in the asymmetric unit. After molecular replacement using an ensemble of known lipases, the resulting phases were used to build the initial model based on the *ORF17* sequence in the Buccaneer software. Several cycles of maximum likelihood refinement and *B*-factor refinement using phenix.refine and REFMAC followed by manual model rebuilding and refitting reduced the *R* values for all data within the resolution range of 38.5 – 1.4 Å. The *R* factor of the present model was 21.9% and the  $R_{\text{free}}$  was 24.6%. The crystallographic statistics for the data collection and structure refinement are summarized in Table 8.

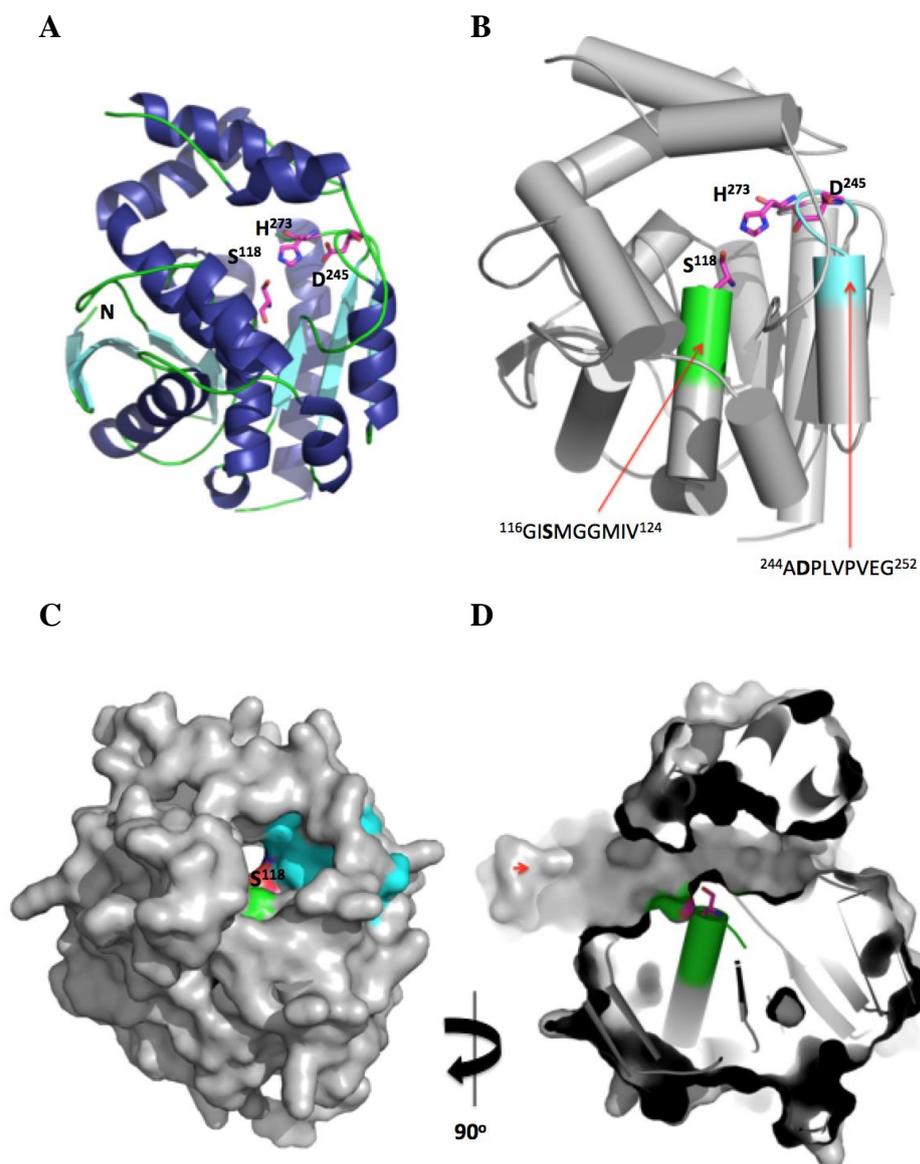
**Table 8 – Data collection and refinement statistics for the *ORF17* protein.**

<i>ORF17</i>	
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (19)
Cell dimensions (Å) <i>a, b, c</i>	43.98 74.95 159.44 90.000 90.000 90.000
Wavelength	0.974710
Resolution (Å)	38.5-1.4
R <sub>merge</sub> (%)	7.6 (76.7)
<i>I</i> / $\sigma$ <i>I</i>	10.17(2.16)
R <sub>meas</sub> (%)	8.2 (88.6)
CC(1/2)	99.9 (51.8)
Redundancy	4.87
<b>Refinement</b>	
Resolution (Å)	38.5 – 1.4
No. of reflections	98406
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	21.9/24.6
Total no. of atoms	4551
<b>r.m.s. deviations</b>	
Bond lengths (Å)	0.006
Bond angles (°)	1.096

#### 4.5.3 Three-dimensional structure of *ORF17*

The three-dimensional structure of *ORF17* consists of an  $\alpha/\beta$ -fold in which the  $\beta$ -sheet is composed of 7  $\beta$ -strands (only strand  $\beta$ 2 is antiparallel) surrounded by thirteen helices (Met36-Ser38, Leu41-Arg50, Ile74-Gln83, Leu93-Leu107, Met109-Asn130, Pro132-Arg134, Asp155-Gln160, Arg170-Tyr184, Asp193-His206, Pro210-Asp223, Val227-Lis232, Val250-Cys259, Cys280-Arg296) (Figure 16A). In this structure, the catalytic Ser118 was located between  $\beta$ 5 and  $\alpha$ 5, while the other two catalytic residues (Asp245 and His273) were found between  $\beta$ 7 and  $\alpha$ 12 and between  $\beta$ 8 and  $\alpha$ 13, respectively (Figure 16A, in pink sticks). The three amino acids of the catalytic triad were located in a large cavity on the opposite side of the N-terminal (Figure 16A). Indeed, the active site motifs found in family V lipases, such as <sup>116</sup>GISMGGM<sup>122</sup> and <sup>245</sup>DPLVPV<sup>250</sup> (Figure 16B, in green and cyan, respectively), were present in the structure of *ORF17*, supporting its assignment to this family.

Interestingly, the structure of *ORF17* shows a tunnel that crosses the protein and allows for the entrance of the substrate up to the active site and the release of the product in the other direction (Figure 16C and 16D). This feature observed in the *ORF17* structure may facilitate the entrance and hydrolysis of a range of different substrates when compared to Est30, which is a highly interesting aspect for biotechnological applications. Activity assays for this protein need to be performed to confirm this hypothesis.



**Figure 16. Three-dimensional structure of *ORF17* protein.** (A) A cartoon representation of the structure with helices in deep blue, beta-sheets in cyan and loops in green. Residues forming the catalytic triad are depicted as pink sticks. (B) The structure of *ORF17* evidencing the two motifs from family V that support the residues of the catalytic triad; (C) A surface view of *ORF17* in gray depicting the detached channel for the substrate and the passage of the product. Ser118 from the catalytic triad is shown in red, and the  $^{116}\text{GISMGMIV}^{124}$  and  $^{244}\text{ADPLVPVEG}^{252}$  motifs are colored in green and cyan, respectively. (D)  $90^\circ$  rotation of (C) evidencing the large channel in the middle of the protein. The red arrow shows the direction of the substrate's entrance.

#### 4.6 Chapter 6: Structural comparison and perspectives

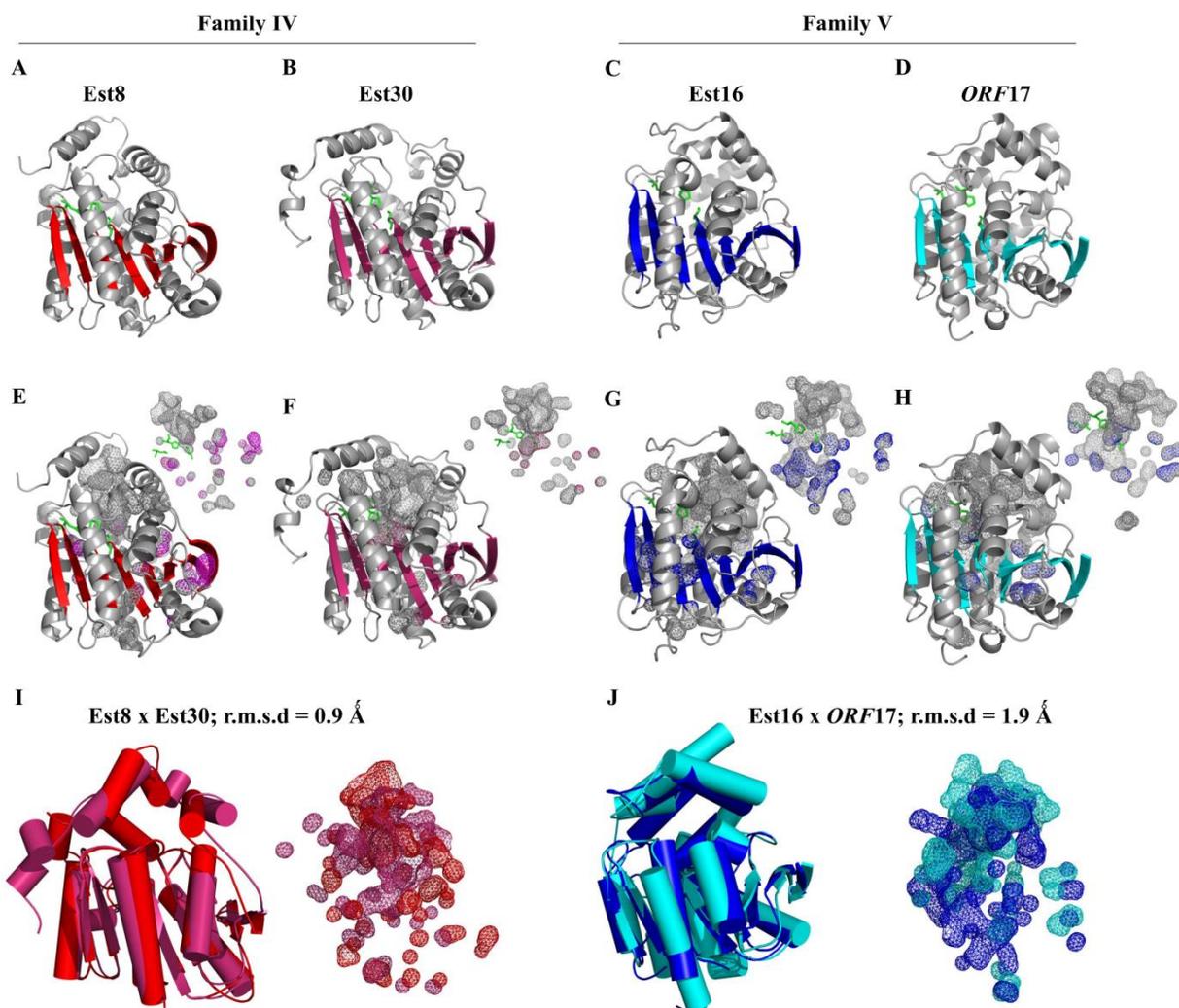
One idea of this Ph.D. project was to characterize the substrate-binding pocket and, when necessary, to establish strategies for rational design in order to target the biotechnological applicability of these enzymes. The characterization of the three-dimensional structures of Est8, Est30 and Est17 (Figure 17A, B and D, respectively) and the structural model of Est16 (Figure 17C) showed that these proteins had interesting features, such as the location of the catalytic triad, the pocket residues, the presence of a channel that gives access to the active site, and conserved motifs (as mentioned in the previous chapters).

Interestingly, by comparing the members of family IV and family V, it was possible to notice a difference in substrate preferences. Est8 showed high activity against *p*-nitrophenyl acetate (C<sub>2</sub>) and a sharp decrease in activity toward *p*-nitrophenyl butyrate (C<sub>4</sub>) (Chapter 3). By contrast, Est30 was able to hydrolyze substrates with acyl chains up to 12 carbons in length (Figure 9A). Although Est8 and Est30 belong to the same family and showed quite identical structures, they showed significant differences in the pocket volume (Figure 17E and F) that affected the range of substrates, indicating that there is diversity in activity even for closely related proteins.

The members of family V exhibited a bigger cavity than the members of family IV (Figure 17E-H). Interesting, the structural analysis supports the result obtained for Est16 in the functional characterization. Namely, this protein was able to hydrolyze substrates with acyl chains up to 16 carbons (Chapter 2). Although the enzymatic assay of *ORF17* has not yet been performed, it is expected that *ORF17* will have a promising applicability due to the newsworthy features of this enzyme. Despite the differences in the cavities of *ORF17* and Est16 (*ORF17* showed a large cavity in its structural analysis), it can be speculated that *ORF17* has a preference for substrates with long chain lengths.

The structural superposition of these proteins again showed that the folding of the members of family IV and family V are highly conserved, as shown by the low r.m.s.d. values between Est8 and Est30 (Figure 17I) and between Est16 and *ORF17* (Figure 17J). It is interesting to note that in most of the cases used for comparison, the  $\alpha/\beta$  domain that contains the secondary structures for sustaining of the catalytic triad is conserved, and the main changes arise from differences in the helical domain that forms the proper active site. Evolutionarily speaking, structural changes in the  $\alpha/\beta$  domain would drastically affect the triad, with unpredictable consequences. On the other hand, changes in the helical domain that

maintain the structure of the triad would guarantee the enzyme's activity but open possibilities for new substrates, which is more interesting for survival.



**Figure 17. Structural comparisons of the family IV lipolytic bacterial enzymes Est8 and Est30 and of the family V lipolytic bacterial enzymes Est16 and ORF17.** The three-dimensional structures or protein models of new members of family IV (A and B) and family V (C and D) are shown. The three-dimensional structures of Est8 (A), Est30 (B) and ORF17 (C) as well as the protein model of Est16 are illustrated in gray cartoons, with the detachment of the beta-sheet from the  $\alpha/\beta$  domain depicted in colors. For comparison, the cavities are shown as internal surfaces in E (Est8), F (Est30), G (Est16) and H (ORF17). The catalytic triads are shown as green sticks. The superpositions of the family IV members depicted in I (with Est8 in red and Est30 in warm pink) and of the family V members depicted in J (with Est16 in blue and ORF17 in cyan) reveal highly conserved three-dimensional structures, with r.m.s.d. values of 0.9 Å and 1.9 Å, respectively. However, there clearly are significant differences in the cavities, as illustrated as the mesh diagrams in the right-side images of I and J.

## 5 CONCLUSIONS

- In this project, it was possible to develop the functional characterization of three new esterases from a microbe consortia specializing in diesel oil degradation;
- Est16 proved to be an esterase capable of hydrolyzing acyl esters with chain lengths of up to 12 carbons. Moreover, this protein is a highly soluble enzyme that is more active in basic pH, thermally stable and DMSO tolerant;
- The structural model of Est16 revealed a highly conserved alpha-beta domain but different substrate pockets than those of patented enzymes, making Est16 a potential candidate for biotechnological applications;
- Est8 is a new alkaline esterase from family IV that displayed a substrate preference for acyl esters with short chain lengths. Interestingly, Est8 showed an optimal temperature at 60 °C and thermal denaturation above 55 °C. In addition, the activity of this protein did not change in the presence of up to 10% DMSO, suggesting its compatibility with organic solvents;
- Through the structure of Est8, it was possible to successfully design the Est8MF mutant. This mutant displayed activity against *p*-nitrophenyl butyrate that was four times higher compared to the wild type;
- Like Est16 and Est8, the Est30 protein is an alkaline esterase. However, Est30 had activity of greater than 60% over a broad temperature range (from 25 °C to 65 °C) and enhanced activity in the presence of DMF;
- In agreement with the functional characterization, the structures of Est30 both in the free form and in a complex with butyrate displayed a tunnel that crossed the protein from one side to another, indicating that long acyl esters could fit in this cavity. The interesting features of Est30 points to its potential for future biotechnological research;
- Although the functional characterization of the *ORF17* protein was not performed, the high-resolution structure of this protein showed a tunnel that crosses the protein. Thus, it is expected that the *ORF17* protein can hydrolyze different acyl esters, making it an attractive candidate for industrial applications.
- Altogether, the data presented in this thesis compile a set of results obtained from the metagenome and structural characterization of new enzymes. These results include the rational evolution of proteins, which reveal the enormous potential of this approach for discovering new potential enzymes with different applications.

## REFERENCES\*

ARPIGNY, J. L.; JAEGER, K-E. Bacterial lipolytic enzymes: classification and properties. **Biochemical Journal**, v. 343, p. 177-183, 1999.

BORNSCHEUER, U. T. Microbial carboxyl esterases: classification, properties and application in biocatalysis. **FEMS Microbiology Reviews**, v. 26, p. 73-81, 2002.

CBDM.T MaBi: The market and business intelligence company reviews the enzyme market. Accessed in November, 2014. [<http://www.biomap.fr/biotech/cbdm-the-market-and-business-intelligence-company-reviews-the-enzyme-market>]

CHOI, Y. J.; LEE, B. H. Culture conditions for the production of esterase from *Lactobacillus casei* CL 96. **Bioprocess and Biosystems Engineering**, v. 24, p. 59-63, 2001.

CHOW, J.; KOVACIC, F.; DALL ANTONIA, Y.; KRAUSS, U.; FERSINI, F.; SCHMEISSER, C.; LAUINGER, B.; BONGEN, P.; PIETRUSZKA, J.; SCHMIDT, M.; MENYES, I.; BORNSCHEUER, U. T.; ECKSTEIN, M.; THUM, O.; LIESE, A.; MUELLER-DIECKMANN, J.; JAEGER, K-E.; STREIT, W.R. The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. **PloS One**, v. 7, p. e47665, 2012.

COUTO, G. H.; GLOGAUER, A.; FAORO, H.; CHUBATSU, L. S.; SOUSA, E. M.; PEDROSA, F. O. Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazil coast. **Genetics and Molecular Research**, v. 9, n. 1, p. 514-523, 2010.

CHU, X.; HE, H.; GUO, C.; SUN, B. Identification of two novel esterases from a marine metagenomic library derived from south China sea. **Applied Microbiology and Biotechnology**, v. 80, p. 615-625, 2008.

DELANO, W. L. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC, New York., 2010.

\* De acordo com: Associação brasileira de Normas Técnicas (ABNT). **NBR 6023**: Informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

EMSLEY, P.; COWTAN, K. Coot: model-building tools for molecular graphics. **Acta Crystallographica Section D**, p. 2126-2132, 2004.

FANG, Z.; LI, J.; WANG, Q.; FANG, W.; PENG, H.; ZHANG, X.; XIAO, Y. A novel esterase from a marine metagenomic library exhibiting salt tolerance ability. **Journal of Microbiology and Biotechnology**, v. 24, p. 771-780, 2014.

FUKUDA, K.; YAMAMOTO, N.; KIYOKAWA, Y.; YANAGIUCHI, T.; WAKAI, Y.; KITAMOTO, K.; INOUE, Y.; KIMURA, A. Balance of activities of alcohol acetyltransferase and esterase in *Saccharomyces cerevisiae* is important for production of isoamyl acetate. **Applied and Environmental Microbiology**, v. 64, n. 10, p. 4076-4078, 1998.

GIULIANI, S.; PIANA, C.; SETTI, L.; HOCHKOEPLER, A.; PIFFERI, P.G.; WILLIAMSON, G.; FAULDS, C.B. Synthesis of pentylferulate by a feruloyl esterase from *Aspergillus niger* using water-in-oil microemulsions. **Biotechnology Letters**, v. 23, p. 325-330, 2001.

Global Industry Analysts. Industrial enzymes - a global multi-client Market research project. GIA: California; 2004.

GLOGAUER, A.; MARTINI, V. P.; FAORO, H.; COUTO, G. H.; MULLER-SANTOS, M.; MONTEIRO, R. A.; MITCHELL D. A.; SOUZA E. M.; PEDROSA, F. O.; KRIEGER, N. Identification and characterization of a new true lipase isolated through metagenomic approach. **Microbial Cell Factories**, v. 10, p. 54-69, 2011.

HANDELSMAN, J.; RONDON, M. R.; BRADY, S. F. ; CLARDY, J.; GOODMAN, R. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chemistry & Biology**, v. 5, n. 10, p. R245-R249, 1998.

HARDEMAN, F.; SJÖLING, S. Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. **FEMS Microbiology Ecology**, v. 59, p. 524-534, 2007.

HENNE, A.; SCHMITZ, R. A.; BOMEKE, M.; GOTTSCHALK, G.; DANIEL, R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. **Applied Environmental Microbiology**, v. 66, p. 3113-3116, 2000.

HORNE, I.; HARCOURT, R.L.; SUTHERLAND, T. D.; RUSSEL, R. J.; OAKESHOTT, J. G. Isolation of *Pseudomonas monteilli* strain with a novel phosphotriesterase. **FEMS Microbiology Letters**, v. 206, p. 51-55, 2002.

HU, Y.; FU, C.; HUANG, Y.; YIN, Y.; CHENG, G.; LEI, F.; LU, N.; LI, J.; ASHFORTH, E. J.; ZHANG, L.; ZHU, B. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. **FEMS Microbiology Ecology**, v. 72, p. 228-237, 2010.

JAEGER, K. E.; DIJKSTRA, B. W.; REETZ, M. T. Bacterial Biocatalysts: Molecular Biology, Three-Dimensional Structures, and Biotechnological Applications of Lipase. **Annual Review Microbiology**, v. 53, p. 315-351, 1999.

JAEGER, K., REETZ, M. T. Microbial lipases form versatile tools for biotechnology. **Trends in Biotechnology**, v. 16, p. 396-403, 1998.

JEON, J. H.; LEE, H. S.; KIM, J. T.; KIM, S. J.; CHOI, S. H.; KANG, S. G.; LEE, J. H. Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment. **Applied Microbiology and Biotechnology**, v.93, p. 623-631, 2012.

JIN, P.; PEI, X.; DU, P.; YIN, X.; XIONG, X.; WU, H.; ZHOU, X.; WANG, Q. Overexpression and characterization of a new organic solvent-tolerant esterase derived from soil metagenomic DNA. **Bioresource Technology**, v. 116, p. 234-240, 2012.

KABSCH, W. XDS. **Acta Crystallographica Section D**, p. 125-132, 2010.

KIM, Y. J.; KWON, E. J.; KIM, S. K.; JEONG, Y. S.; KIM, J.; YUN, H. D.; KIM, H. Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. **Biochemical and Biophysical Research Communications**, n. 393, p. 45-49, 2010.

KOBLITZ, M. G. B.; PASTORE, G. M. Purificação parcial, por dois diferentes métodos cromatográficos, da lipase produzida por *Rhizopus* sp. **Ciência e Tecnologia de Alimentos**, v. 24, n. 2, p. 287-292, 2004.

LEE, S. W.; WON, K.; LIM, H. K.; KIM, J. C.; CHOI, G. J.; CHO, K. Y. Screening for novel lipolytic enzymes from uncultured soil microorganisms. **Applied Microbiology and Biotechnology**, v. 65, p. 720-726, 2004.

[LONG, C. **Biochemists handbook**. London: Redwood, 1971. 1192?p.]

LORENZ, P.; ECK, J. Metagenomics and industrial applications. **Nature**, v. 3, p. 510-516, 2005.

MURSHUDOV, G. N.; SKUBÁK, P.; LEBEDEV, A. A.; PANNU, N. S.; STEINER, R. A.; NICHOLLS, R. A.; WINN, M. D.; LONG, F.; VAGIN, A. A. REFMAC5 for the Refinement of Macromolecular Crystal Structures. **Acta Crystallographica Section D**, p. 355-367, 2011.

NARDINI, M.; DIJKSTRA, B. W.  $\alpha/\beta$  Hydrolase fold enzymes: the family keeps growing. **Current opinion in Structural Biology**, v. 9, p. 732-737, 1999.

OLLIS, D. L.; CHEAH, E.; CYGLER, M.; DIJKSTRA, B.; FROLOW, F.; FRANKEN, S. M.; HAREL, M.; REMINGTON, S. J.; SILMAN, I.; SCHRAG, J. The alpha/beta hydrolase fold. **Protein Engineering, Design and Selection**, v. 5, n. 3, p. 197-211, 1992.

PENG, Q.; WANG, X.; SHANG, M.; HUANG, J.; GUAN, G.; LI, Y.; SHI, B. Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. **Microbial Cell Factories**, v. 13, n. 1, 2014.

PERÄNEN J., RIKKONEN M., HYVÖNEN M., KÄÄRIÄINEN L. T7 vectors with modified T7 lac promoter for expression of proteins in *Escherichia coli*. **Analytical Biochemistry**, p. 371-373, 1996.

QUAX, W.J.; BROEKHUIZEN, C.P. Development of a new Bacillus carboxyl esterase for use in the resolution of chiral drugs. **Applied Microbiology and Biotechnology**, v. 41, n. 4, p. 425-431, 1994.

RANJAN R.; GROVER, A.; KAPARDAR, R. K.; SHARMA, R. Isolation of novel lipolytic genes from uncultured bacteria of pond water. **Biochemical and Biophysical Research Communications**, n. 335, p. 57-65, 2005.

RENGACHARI, S.; ASCHAUER, P.; SCHITTMAYER, M.; MAYER, N.; GRUBER, K.; BREINBAUER, R.; BIRNER-GRUENBERGER, R.; DREVENY, I.; OBERER, M. Conformational plasticity and ligand binding of bacterial monoacylglycerol lipase. **The Journal of Biological Chemistry**, v. 288, n. 43, p. 31093-31104, 2013.

ROH, C.; VILLATTE, F. Isolation of a low-temperature adapted lipolytic enzyme from uncultivated micro-organism. **Journal of Applied Microbiology**, n. 105, p. 116-123, 2008.

RONDON, M. R.; AUGUST, P. R.; BETTERMANN, A. D.; BRADY, S. F.; GROSSMAN, T. H.; LILES, M. R.; LOIACONO, K. A.; LYNCH, B. A.; MACNEIL, I. A.; MINOR, C.; TIONG, C. L.; GILMAN, M.; OSBURNE, M. S.; CLARDY, J.; HANDELSMAN, J.; GOODMAN, R. M. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. **Applied Environmental Microbiology**, v. 66, n. 6, p.2541-2547, 2000.

SAXENA, R. K.; SHEORAN, A.; GIRI, B.; DAVIDSON, W. S. Purification strategies for microbial lipases. **Journal of Microbiological Methods**, v. 52, n. 1, p. 1-18, 2003.

SCHÜETTELKOPF, A. W.; VAN AALTEN D. M. F. *PRODRG*: a tool for high-throughput crystallography of protein-ligand complexes. **Acta Crystallographica Section D**, p. 1355-1363, 2004.

SEO, S.; LEE, Y-S.; YOON, S-H; KIM, S-J.; CHO, J. Y.; HAHN, B-S.; KOO, B-S.; LEE, C-M. Characterization of a novel cold-active esterase isolated from swamp sediment metagenome. **World Journal of Microbiology and Biotechnology**, v. 30, p. 879-886, 2014.

SILVEIRA, E. L.; PEREIRA, R. M.; SCAQUITTO, D. C.; PEDRINHO, E. A. N.; VAL-MORAES, S. P.; WICKERT, E.; CARARETO-ALVES, L. M.; LEMOS, E. G. M. Bacterial diversity of soil under eucalyptus assessed by 16S rDNA sequencing analysis. **Pesquisa Agropecuária Brasileira**, v. 41, n. 10, p. 1507-1516, 2006.

STEELE, H.; STREIT, W. R. Metagenomics: Advances in ecology and biotechnology. **FEMS Microbiology Letters**, v. 247, p. 105-111, 2005.

UPPENBERG, J.; OHRNER, N.; NORIN, M.; HULT, K.; KLEYWEGT, G. J.; PATKAR, S.; WAAGEN, V.; ANTHONSEN, T.; JONES, T. A. Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. **Biochemistry**, p. 16838-16851, 1995.

VAKHLU, J.; KOUR, A. Yeast lipases: enzyme purification, biochemical properties and gene cloning. **Journal of Biotechnology**, v. 9, n. 1, p. 69-85, 2006.

WU, C.; SUN, B. Identification of novel esterase from metagenomic library of Yangtze River. **Journal of Microbiology and Biotechnology**, v. 19, n. 2, p. 187-193, 2009.

**Annex A - List of patented esterases/lipases used in building the phylogenetic tree containing the four enzymes studied in this Ph.D. project.** Patented enzymes were obtained in a search of the literature and from the Questel website using the Orbit database. The sequences of these patents were obtained at the GenomeQuest site (<http://www.genomequest.com/>).

Reference code	Microorganism	Protein function	Protein identification/ Patent number
Unc_US20050153404_pharma	Unculturable strain	Lipase - preparing of enantiomeric enriched esters and alcohols	US20050153404
Psa_WO2010134035_food	<i>Pelomonas saccharophila</i>	Lipase - combined enzymes, preparing bread and baking	WO2010134035
Bsp_WO2008032007_biofuel	<i>Bacillus</i> sp.	Lipase - biodiesel production	WO2008032007
Sgr_WO2011150157_detergent	<i>Streptomyces griseus</i>	Lipase - highly suitable as a detergent	WO2011150157
Mth_WO2012078741_industrial	<i>Myceliophthora thermophila</i>	Esterases - industrial food, bioremediation, biorefining, detergents	WO2012078741
Pal_WO9530744_detergent	<i>Pseudomonas alcaligenes</i>	Lipase - improved resistance to surfactants	WO9530744
Rsp_WO2007128496_chemicals	<i>Rhodococcus</i> sp.	Esterase - production of fine chemicals	WO2007128496
Unk_WO2007128441_chemicals	Unknown	Esterases - production of fine chemicals	WO2007128441
Fox_US20030180418_food	<i>Fusarium oxysporum</i>	Lipase - dough or baked products	US20030180418
Unk_WO2009106575_food	Unknown strain	Lipase - dairy or baking industry	WO2009106675
Mgr_WO2007096201_food	<i>Magnaporthe griseae</i>	Lipase - baking, degumming, food emulsifiers	WO2007096201
Tla_WO2006136159_pharma	<i>Thermomyces lanuginosus</i>	Lipase - treatment of human disorders	WO2006136159
Sri_WO2010065455_detergent	<i>Streptomyces grimosus</i>	Lipase - removal of oily stains from fabric	WO2010065455
Ssp_WO2006008653_food	<i>Streptomyces</i> sp.	Lipase - hydrolysis of glycolipids and phospholipids	WO2006008653
Dgr_WO2004064537_food	<i>Digitalis grandiflora</i>	Lipase - food emulsifiers	WO2004064537
Lre_US20100216212_pharma	<i>Lactobacillus reuteri</i>	Lipase - anti-obesity activity	US20100216212
Csp_CN101957838_bioremediation	<i>Culex</i> sp.	Esterases - bioremediation of organic phosphorus	CN101957838
Lcu_US5843758_bioremediation	<i>Lucilia cuprina</i>	Esterases - bioremediation of organic phosphate and carbamate	US5843758

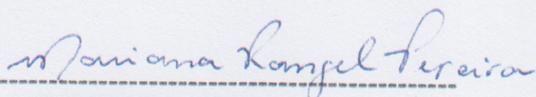
**Annex B - The co-authors' authorization to use the Est16 manuscript as a chapter in this Ph.D. thesis.**

February 09<sup>th</sup>, 2015

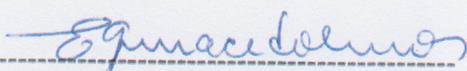
Term of agreement

We reviewed this document to authorize the inclusion of the Est16 manuscript as a chapter in Mariana Rangel Pereira's thesis. The experiments of this manuscript, which is titled **"Est16, a New Esterase Isolated from Metagenomic Library of a Microbe Consortium Specializing in Diesel Oil Degradation"**, were carried out by Mariana under the supervision of Dra. Eliana G. de Macedo Lemos and Dra. Andrea Balan. The screening for lipolytic activity and the kinetics analysis were performed in collaboration with Thaís C. Maester and Gustavo F. Mercaldi, respectively.

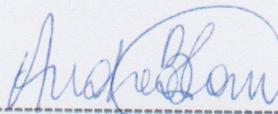
To be true, we signed,



Mariana Rangel Pereira  
Ph.D student



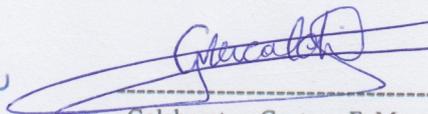
Prof. Dra. Eliana G. de Macedo  
Department of Technology,  
São Paulo State University,  
Jaboticabal, São Paulo State, Brazil.



Prof. Dra. Andrea Balan  
Department of Microbiology,  
Biomedical Sciences Institute II,  
University of São Paulo,  
São Paulo, São Paulo State, Brazil.



Colaborator: Thaís Carvalho Maester  
Ph.D. student



Colaborator: Gustavo F. Mercaldi  
Ph.D. student

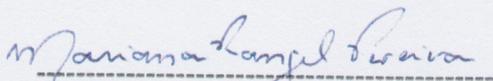
**Annex C - The co-authors' authorization to use the Est8 manuscript as a chapter in this Ph.D. thesis.**

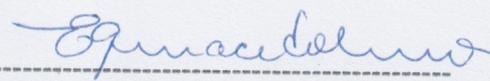
February 09<sup>th</sup>, 2015

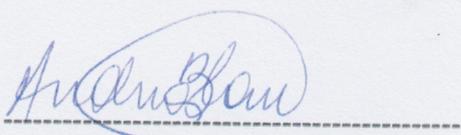
Term of agreement

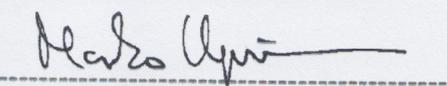
We reviewed this document to authorize the inclusion of the Est8 manuscript as a chapter in Mariana Rangel Pereira's thesis. The experiments of this manuscript, which is titled "**High-resolution of Est8 metagenome-derived extends the esterase's diversity**", were carried out by Mariana under the supervision of Dra. Eliana G. de Macedo Lemos, Dra. Andrea Balan and Dr. Marko Hyvonen. The screening for lipolytic activity and the kinetics analysis were performed in collaboration with Thaís C. Maester and Gustavo F. Mercaldi, respectively.

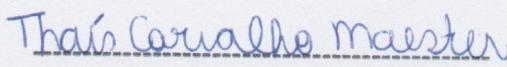
To be true, we signed,

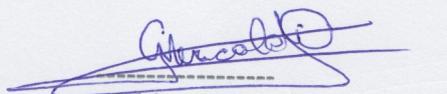
  
-----  
Mariana Rangel Pereira  
Ph.D. student

  
-----  
Prof. Dra. Eliana G. de Macedo  
Department of Technology,  
São Paulo State University,  
Jaboticabal, São Paulo State, Brazil.

  
-----  
Prof. Dra. Andrea Balan  
Department of Microbiology,  
Biomedical Sciences Institute II,  
University of São Paulo,  
São Paulo, São Paulo State, Brazil.

  
-----  
Prof. Dr. Marko Hyvonen  
Department of Biochemistry  
University of Cambridge, UK.

  
-----  
Colaborator: Thaís Carvalho Maester  
Ph.D. student

  
-----  
Colaborator: Gustavo F. Mercaldi  
Ph.D. student