

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Departamento de Genética e Evolução
Laboratório de Biologia Molecular

CÉLIO DIAS SANTOS JÚNIOR

**CONVERSÃO DE UMA HISTIDINA AMÔNIA LIASE
PARA FENILALANINA AMÔNIA LIASE POR MEIO DE
BIOLOGIA SINTÉTICA**

SÃO CARLOS - SP
2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Departamento de Genética e Evolução
Laboratório de Biologia Molecular

CÉLIO DIAS SANTOS JÚNIOR

**CONVERSÃO DE UMA HISTIDINA AMÔNIA LIASE
PARA FENILALANINA AMÔNIA LIASE POR MEIO DE
BIOLOGIA SINTÉTICA**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos para obtenção do título de Mestre em Genética Evolutiva e Biologia Molecular.

Orientador: Prof. Dr. Flávio Henrique Silva

SÃO CARLOS - SP
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

S237c Santos Júnior, Célio Dias
Conversão de uma histidina amônia liase para
fenilalanina amônia liase por meio de biologia
sintética / Célio Dias Santos Júnior. -- São Carlos
: UFSCar, 2016.
121 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2016.

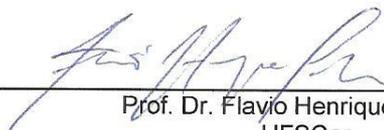
1. Fenilalanina amônia-liase. 2. Histidina amônia-
liase. 3. Lago Poraquê. 4. Engenharia de enzimas. 5.
Produção recombinante. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia
Molecular

Folha de Aprovação

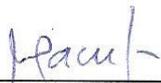
Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Célio Dias Santos Júnior, realizada em 24/02/2016:



Prof. Dr. Flávio Henrique da Silva
UFSCar



Prof. Dr. Felipe Roberti Teixeira
UFSCar



Prof. Dr. Maurício Bacci Júnior
UNESP

DEDICATÓRIA

Dedico este trabalho aos meus pais, Célio e Sandra, pelo amor incondicional, incentivo e apoio. Dedico também aos meus avós e amigos. Essa conquista não seria possível sem a minha família biológica e de alma.

AGRADECIMENTOS

À minha família, especialmente meus avós, pais e irmã, pelo apoio, paciência, confiança e incentivo.

Ao Prof. Dr. Flavio Henrique Silva pela orientação e pelos ensinamentos ao longo desses anos.

À Profa. Dra. Andrea Soares Costa Fuentes, por todo o aprendizado que me proporcionou e por sua disposição em ajudar sempre que precisei.

Ao Prof. Dr. Francis Nunes pela paciência, cortesia, ensinamentos e confiança ao longo desses anos.

Ao Prof. Dr. Júlio César Borges e seus alunos pela cortesia e ensinamentos os quais me auxiliaram a concluir o presente trabalho.

Ao Dr. Luciano Kishi pela sua colaboração na montagem e produção do banco de dados utilizado neste estudo.

Aos Doutorandos Danyelle Toyama, Adelita Santiago, Vanessa Schneider e Darlan Nakayama, pela valiosa amizade, pelos ensinamentos e pela ajuda prestados.

Aos Mestrandos Bruno Lima e Augusto Balula, pela convivência, ajuda e experiência compartilhada, além das descontraídas “horas do café”.

À Técnica Elizabeth Baraldi por todo o suporte e auxílio no laboratório, pela confiança e amizade.

Às amigas Natália Melquie, Thaís Ferreira Isabel, Dávilla Martins, Rejane Monte, Ana Carolina Cleto, Lorena Cruz, Isabella Kitano, Izabela Watanabe, Nayara Costa, Maiza Carvalho e a todas as outras amigas as quais não foi possível citar, pela maravilhosa e animada companhia durante esses anos, nas horas boas e ruins. Pelo incentivo e ajuda e por tornar meu dia-a-dia mais alegre.

Aos amigos Rodolfo Cruz Jr., Jordi Brines, Marcelo Castro, Denis Prudêncio, Rodrigo Alves, Michael Zanin e a todos os outros amigos os quais não foi possível citar, pela

convivência, auxílio, esforço, perseverança e apoio nos momentos de desespero e alegria.

Aos amigos de Araxá - MG e Uberlândia - MG, que apesar da distância sempre me incentivaram.

Aos amigos são-carlenses, que tive o prazer de conhecer e se tornaram especiais para mim.

Ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular pelo auxílio e suporte.

Ao CNPq pela bolsa de estudos.

RESUMO

A fenilalanina amônia liase (PAL) catalisa a desaminação não oxidativa de L-Phe. Ela é uma enzima importante na produção de misturas enântio-puras de fenilalanina. Além de ser importante no metabolismo secundário vegetal, a PAL também tem sido utilizada tanto para aplicações industriais quanto no tratamento de leucemia e fenilcetonúria. As PAL apresentam alta similaridade com as histidina amônia liases (HAL) e acredita-se que as PAL se originaram das HAL. Utilizamos sequências das extremidades de uma HAL, previamente obtidas do metagenoma do Lago Poraquê (03°57' S e 63°10' W) para produção de uma enzima engenheirada. A HAL montada a partir delas teve seu núcleo catalítico trocado por um que foi desenhado a partir de sequências de PAL previamente descritas. Desta forma visamos deslocar as atividades da HAL tornando-a ativa em presença de fenilalanina por meio de um protocolo simplificado. As nossas análises filogenéticas revelam uma divisão clara entre PALs vegetais e fúngicas, e mostram que a recombinação teve um papel fundamental na separação destes grupos. Além disto, também observamos que o núcleo catalítico da enzima é conservado, com relação às extremidades. A enzima recombinante, mPAL_c1, foi produzida sob forma insolúvel e renovelada *in vitro*. A mPAL_c1 tendo como substrato a L-Phe, possui uma T_{opt} de 30°C num pH de 7,5 com 2mM de $MnCl_2$ e um K_M de 55 μ M, $V_{Máx}$ de 15mU e K_{cat}/K_M de 0,01633 $mM^{-1}s^{-1}$. A atividade da mPAL_c1 se mostrou cerca de 30% da atividade da PAL comercial de *Rhodotorula toruloides* a 2 mM de L-Phe. A ativação pelo substrato foi de quase 5 vezes quando a concentração de L-Phe foi elevada de 2mM até 5mM. Os resultados sugerem um resquício de atividade de histidina amônia-liase. As baixas taxas catalíticas se justificam devido à agregação e *misfolding* proteicos, detectados pela cromatografia de exclusão molecular e pelo dicróismo circular. Por fim, o protocolo estabelecido neste estudo se mostrou útil para a engenharia de proteínas.

PALAVRAS CHAVE: fenilalanina amônia-liase, histidina amônia-liase, lago Poraquê, engenharia de enzimas, proteína heteróloga.

ABSTRACT

Phenylalanine ammonia lyase (PAL) catalyzes the non-oxidative deamination of L-Phe. It is an important enzyme in the production of enantio-pure mixtures of phenylalanine. Besides being important in plant secondary metabolism, PAL has also been used both for industrial applications as in the treatment of leukemia and phenylketonuria. The PAL have high similarity with histidine ammonia lyase (HAL) and it is believed that the PAL originated from HAL. Use of HAL termini sequences, previously obtained from the metagenome of Poraquê Lake (03°57' S, 63°10' W) to produce an engineered enzyme. The HAL assembled from them had its catalytic core exchanged by one which was designed from the previously described PAL sequences. Thus, We aim to shift the HAL activity making it active in the presence of phenylalanine through an easy protocol. Our phylogenetic analysis shows a clear division between plant and fungal PALs, and showed that recombination played a key role in the separation of these groups. In addition, We also observed that PAL catalytic core is conserved despite positive selection operating in protein termini. The recombinant enzyme, mPAL_c1, was produced in insoluble form and was refolded in vitro. mPAL_c1, when using L-Phe as substrate, has a $T_{OPT.}$ of 30°C (at pH 7.5 with 2mM $MnCl_2$), K_M of 55 μ M and K_{cat}/K_M of 0.01633 $mM^{-1}s^{-1}$. The activity of mPAL_c1 showed about 30% of activity of commercial PAL from *Rhodotorula toruloides* at 2 mM L-Phe. Activation of mPAL_c1 was tested through substrate concentration rising from 2 mM to 5 mM of L-Phe, and its activity was almost 5 times higher. The results still suggest a histidine ammonia lyase remnant activity. The low catalytic rates are justified due to protein aggregation and misfolding, detected by size exclusion chromatography and by circular dichroism. In summary, our protocol has proved to be useful in protein design.

KEY WORDS: phenylalanine ammonia-lyase, histidine ammonia-lyase, Poraquê Lake, enzymes engineering, recombinant protein.

LISTA DE TABELAS

Tabela 1. Principais características da fenilalanina amônia liase de diferentes fontes....	2
Tabela 2. Divergência evolutiva das isoformas de AtPAL1-3 e diferentes sequências de PAL.....	12
Tabela 3. Principais espécies de fungos e suas respectivas características da PAL.....	13
Tabela 4. Sequências ortólogas de PAL e Actina utilizadas neste estudo.....	25
Tabela 5. Correlação entre as características da PAL.....	29
Tabela 6. Agrupamento diferencial da PAL por meio de diferentes abordagens.....	35
Tabela 7. Estimativas de divergência evolutiva entre grupos de sequências	40
Tabela 8. Resultados do teste de relógios moleculares utilizando o método de máxima verossimilhança (ML).....	44
Tabela 9. Eventos recombinatoriais nas sequências de PAL.....	51
Tabela 10. Principais proteínas padrão utilizadas no experimento de calibração.	62
Tabela 11. Constantes de reação para cálculo da atividade em diferentes condições ...	65
Tabela 12. Comparação dos resíduos em torno dos sítios ativos da PAL.....	69
Tabela 13. Análise dos pockets da estrutura mPAL_c1.	72
Tabela 14. Resumo das técnicas de refolding utilizadas na mPAL_c1	78
Tabela 15. Principais características hidrodinâmicas da mPAL_c1 calculadas pelos dados da cromatografia de exclusão molecular.....	82
Tabela 16. Estrutura secundária do agregado mPAL_c1 por dicróismo circular sob deconvolução em diferentes softwares.	87
Tabela 17. Efeito de diferentes cátions divalentes sobre PALs isoladas de diferentes fontes	88

LISTA DE FIGURAS

Figura 1. Estrutura quaternária da PAL.....	4
Figura 2. Alinhamento evidenciando as extensões típicas da PAL.....	4
Figura 3. Mecanismo de reação da PAL	5
Figura 4. Principais compostos derivados da PAL.....	7
Figura 5. Via bioquímica do Cinamoil-coA.....	8
Figura 6. Vias de indução de repressão da atividade de PAL nas plantas.....	10
Figura 7. Cenário evolutivo para surgimento da PAL vegetal e fúngica	16
Figura 8. Filogenia das enzimas PAL, TAL e HAL.....	17
Figura 9. Desvios do uso de aminoácidos detectados no set de sequências da proteína PAL.....	34
Figura 10. Análise de agrupamento de táxons de acordo com as principais características bioquímicas da PAL.....	36
Figura 11. Variabilidade da proteína PAL	39
Figura 12. Reconstrução filogenética da história evolutiva da PAL pelo método ML .	42
Figura 13. Gráfico de <i>Splits</i> da PAL completa.....	45
Figura 14. Seleção positiva operando em cada códon da sequência de PAL.....	47
Figura 15. Eventos recombinantes envolvidos na evolução da PAL	48
Figura 16. Resumo gráfico da construção de uma PAL híbrida.....	52
Figura 17. Esquema de purificação da mPAL_c1 por meio de <i>refolding</i> por diluição rápida	60
Figura 18. Montagem da mPAL_c1 mostrando regiões onde o <i>core</i> conservado foi inserido entre as ORFs da HAL que o recebeu.	68

Figura 19. Estrutura mPAL_c1	71
Figura 20. Alinhamento estrutural da mPAL_c1 (em branco) e 1w27_A (em azul)	72
Figura 21. Análise filogenética com as PAL parentais do <i>core</i> catalítico da mPAL_c1	73
Figura 22. Expressão (acima) e <i>Western Blot</i> (abaixo) da mPAL_c1 recombinante ...	75
Figura 23. IMAC em ambiente desnaturante para purificação da mPAL_c1	76
Figura 24. SDS-PAGE do processo de purificação da mPAL_c1	77
Figura 25. Análise da reconstituição do cofator MIO (4-metilidenoimidazol-5-ona) depois do processo de <i>refolding</i>	79
Figura 26. Cromatografia de exclusão molecular da mPAL_c1 renaturada	81
Figura 27. Ensaio da emissão de fluorescência do Trp	84
Figura 28. Dicroísmo circular do agregado da mPAL_c1	86
Figura 29. Ensaio de atividade da mPAL_c1	89
Figura 30. Constantes cinéticas comparadas do agregado mPAL_c1 com as PAL do banco de dados BRENDA-ENZYME	91
Figura 31. Ativação da mPAL_c1 por substrato e comparação com a enzima comercial de <i>R. glutinis</i>	93
Figura 32. Efeito do amônio na mPAL_c1 com diferentes substratos.	93

LISTA DE ABREVIACOES E SIGLAS

(-)/(+)	Relao de cargas dos aminocidos de uma protena
(+)F	Frequncias discretas
(+)G	Distribuio gama
(+)I	Stios invariantes
$[\theta]_{MRW(\lambda)}$	Elipsidade residual mdia
°C	Graus Celsius
μg	Microgramas
μL	Microlitros
μM	Micromolar
μm	Micrmetros
2,4D	cido diclorofenxiactico
aa	Aminocidos
Abs	Absoro
AI	ndice aliftico
AICc	Crterio informativo de Akaike corrigido
ANOVA-1F	Anlise de varincia de 1 fator
Asn	Asparagina
ATP	Adenosina Trifosfato
AtPAL	Isoforma da fenilalanina amnia-liase de <i>Arabidopsis thaliana</i>
BEP	Clado das gramneas que rene: <i>Bambusoideae</i> , <i>Ehrhartoideae</i> e <i>Pooideae</i>
BIC	Crterio informativo Bayesiano
BKP	Ponto de ruptura para recombinao (<i>Breaking point</i>)
Blast	<i>Basic local alignment search tool</i>
BRENDA	<i>Braunschweig enzyme database</i>
BSA	Soroalbumina bovina
C	Carbono
<i>c</i>	Concentrao de protena em mg/mL
CD	Dicrosmo circular
cDNA	DNA complementar

cm	Centímetros
coA	Coenzima A
D+E	Conteúdo de aminoácidos negativos (aspártico + glutâmico)
Da	Dalton (g/mol)
dmol	Decimol
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeos trifosfato
D-Phe	Isômero dextrógiro da fenilalanina
DTT	1,4-ditiotreitol
D-Tyr	Isômero dextrógiro da tirosina
E.C.	<i>Enzyme Commission number</i>
g	Gramas
GARD	<i>Genetic algorithm for recombination detection,</i>
GC%	Conteúdo molar percentual de guanidina e citosina
Gln	Glutamina
GRAVY	Coefficiente médio de hidrofobicidade
h	Horas
H	Hidrogênio
HAL	Histidina amônia-liase
HCl	Ácido clorídrico
HMM	Modelo de cadeias ocultas de Markov
HT voltage	Voltagem dos tubos fotomultiplicadores do Dinodo (<i>Dynode Voltage</i>)
IAA	Ácido indol-3-acético
IMAC	Cromatografia de afinidade à metal imobilizado
ININ	Índice de Instabilidade
K	Graus Kelvin ou Lisina
KCl	Cloreto de potássio
K _{av}	Coefficiente de partição
K _{cat}	Número de <i>Turnover</i>
kDa	Quilodaltons
KEGG	Enciclopédia de Genes e Genomas de Kioto
Kpb	Quilopares de base
KH	Teste de Kishino-Hasegawa
K _M	Constante de Michaelis-Menten

kV	Quilovolts
L	Litros
<i>l</i>	Comprimento do caminho óptico
LB	Luria-bertani
L-Cys	Isômero levógiro da cisteína
LG	Matriz de substituição de Le e Gascuel (2008)
L-His	Isômero levógiro da histidina
LnL	Logaritmo natural da Verossimilhança
L-Phe	Isômero levógiro da fenilalanina
L-Tyr	Isômero levógiro da tirosina
M	Molar (mol/L)
mA	Miliampere
MaxChi	Teste de Qui-quadrado máximo
mdeg	Miligráus de absorbtividade da luz circularmente polarizada
MEGA	<i>Molecular Evolutionary Genetics Analysis Software</i>
MetaCyc	<i>Metabolic Pathways From all Domains of Life Webserver</i>
mg	Miligramas
min	Minutos
MIO	4-metilideno imidazol-5-ona
ML	Algoritmo de verossimilhança máxima (<i>Maximum likelihood</i>)
mL	Mililitros
mM	Milimolar
mm	Milímetros
MnCl ₂	Cloreto de manganês
MgCl ₂	Cloreto de magnésio
mPAL_c1	Metafenilalanina amônia liase consenso 1
MUBs	Desvio do uso de monômeros (<i>Monomer Usage Bias</i>)
MW	Massa molecular (<i>Molecular weight</i>)
MWCO	Ponto de corte por massa molecular (<i>Molecular Weight Cut-off</i>)
N	Nitrogênio; Asparagina
NaBH ₄	Borohidreto de sódio
NaB ³ H ₄	Borohidreto de sódio tritiado
NaCl	Cloreto de Sódio
NADH	Nicotinamida adenina dinucleotídeo

NaH ₂ PO ₄	Fosfato de sódio monobásico
N/C	Razão entre conteúdos atômicos de nitrogênio e carbono
NH ₄ ⁺	Amônio
NH ₄ Cl	Cloreto de amônio
Ni ²⁺ -NTA	Resina de ácido nitriloacético carregado com íons de níquel
NJ	<i>Neighbor-joining</i>
nm	Nanômetros
nM	Nanomolar
NNI	<i>Nearest-neighbor-interchange</i>
NRes.	Número de resíduos de uma proteína
O	Oxigênio
OD	Densidade óptica
ORF	Fase de leitura aberta (<i>Open reading frame</i>)
OTU	Unidade taxonômica operacional (<i>Operational Taxonomic Unit</i>)
PACMAD	Clado das gramíneas que reúne: <i>Panicoideae</i> , <i>Arundinoideae</i> , <i>Chloridoideae</i> , <i>Micrairoideae</i> , <i>Aristidoideae</i> e <i>Danthonioideae</i>
PAL	Fenilalanina amônia-liase
PAL_Rho	Fenilalanina amônia-liase comercial de <i>Rhodotorula glutinis</i>
pb	Pares de base
PCR	Reação em cadeia pela polimerase
PDB	<i>RCSB Protein Data Bank</i> (incluindo formato de arquivos)
pH	Potencial hidrogeniônico
pI	Ponto isoelétrico
pmol	Picomoles
PMT Volt.	Voltagem dos tubos fotomultiplicadores (<i>Photomultiplier tubes voltage</i>)
PTAL	Fenilalanina/Tirosina amônia-liase
PVDF	Fluoreto de polivinilideno
R+K	Conteúdo de aminoácidos positivos (arginina + lisina)
r	Coefficiente de correlação de Spearman
Rd	Diluição rápida
RDP	Programa de detecção de recombinação (<i>Recombination Detection Program</i>)
Rh	Raio hidrodinâmico (Raio de Stokes)
Rh/R ₀	Razão friccional

R ₀	Raio hidrodinâmico de uma esfera de massa idêntica à da proteína, no caso se ela fosse uma esfera completamente lisa e compacta.
RNA	Ácido ribonucleico
ROS	Espécie reativa de oxigênio (<i>Reactive Oxygen Species</i>)
rpm	Rotações por minuto
s	Segundos
S	Enxofre
SD	Desvio padrão (<i>Standard Deviation</i>)
SDS-PAGE	Eletroforese em gel de SDS-Poliacrilamida
TAL	Tirosina amônia-liase
Taq	<i>Thermus aquaticus</i>
t-CA	Ácido trans-cinâmico
t-CM	Trans-cinamilo
Thr	Treonina
T _{ÓPT.}	Temperatura ótima
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Trp	Triptofano
U	Atividade enzimática em unidades internacionais (1 μmol/min/L)
U.A.	Unidade arbitrária (seja enzimática ou de fluorescência)
UPGMA	Método de agrupamento par-a-par não ponderado com média aritmética
UPGMB	Método de agrupamento par-a-par não ponderado combinado
UV-Vis	Região do espectro da radiação que abrange dos comprimentos da luz ultravioleta à da região visível
V	Volts
V _e	Volume de Eluição (<i>Elution volume</i>)
V _f :V _i	Do Volume final para o volume inicial
V _i :V _f	Do Volume inicial para o volume final
V _{Máx}	Velocidade máxima
V _o	Volume "morto" (<i>Void volume</i>)
V _T	Volume total (<i>Total volume</i>)
WB	<i>Western blot</i>
xg	Força Centrífuga Relativa em vezes da força gravitacional da Terra
ε	Coefficiente de absorbtividade molar (M ⁻¹ cm ⁻¹)
λ	Comprimento de onda

SUMÁRIO

DEDICATÓRIA	i
AGRADECIMENTOS	ii
RESUMO	iv
ABSTRACT	v
LISTA DE TABELAS	vi
LISTA DE FIGURAS	vii
LISTA DE ABREVIÇÕES E SIGLAS	ix
Capítulo 1. REVISÃO TEÓRICA	1
<i>Bioquímica, Mecanismos e Estrutura</i>	1
<i>Principais Vias</i>	6
<i>Indução e Repressão: uma nova perspectiva sobre dados antigos</i>	8
<i>Diversidade e Evolução: estado da arte</i>	11
<i>Aplicações biotecnológicas</i>	19
<i>A biologia sintética e design de proteínas</i>	20
<i>Objetivos e Justificativas</i>	21
Capítulo 2. INSIGHTS SOBRE A EVOLUÇÃO DA FENILALANINA AMÔNIA LIASE (PAL): EVENTOS DE RECOMBINAÇÃO PODERIAM TRAZER DIVERSIDADE À SUA HISTÓRIA?	23
MATERIAL E MÉTODOS	23
RESULTADOS E DISCUSSÃO	28
<i>Características da Proteína</i>	28
<i>Inferências filogenéticas</i>	37
<i>Evolução molecular</i>	43
Capítulo 3. DESIGN DE UMA CONSTRUÇÃO QUIMÉRICA PARA SUBSTITUIÇÃO DE SUBSTRATO: CONVERSÃO DE UMA HAL EM PAL ...	52
MATERIAL E MÉTODOS	52
<i>Design da proteína mPAL_c1</i>	52
<i>Testes de confirmação da sequência mPAL_c1</i>	55
<i>Manipulações com a sequência mPAL_c1</i>	56
<i>Expressão recombinante de mPAL_c1</i>	58
<i>Determinação da concentração de mPAL_c1 purificada</i>	61
<i>Medição de estado oligomérico</i>	62

<i>Medição da restituição do co-fator MIO por varredura espectroscópica</i>	63
<i>Medição do $\lambda_{M\acute{a}x}$ da emissão de triptofanos</i>	63
<i>Dicroísmo circular</i>	63
<i>Medição de atividade da mPAL_c1 purificada</i>	64
<i>Efeitos da temperatura, pH e inibição por amônia</i>	66
<i>Determinação das constantes cinéticas da enzima (K_M, $V_{M\acute{a}x}$. e K_{cat})</i>	66
RESULTADOS E DISCUSSÃO	67
<i>Análise bioinformática</i>	67
<i>Análise filogenética</i>	73
<i>Expressão heteróloga</i>	74
<i>Qualidade estrutural da mPAL_c1</i>	78
<i>Atividade enzimática do agregado mPAL_c1</i>	88
CONCLUSÕES	94
REFERÊNCIAS	95
ANEXO I - Via da Fenilalanina (KEGG Pathway 00360)	112
ANEXO II - Biossíntese de Fenilpropanóides (KEGG Pathway 00940)	113
ANEXO III - Curvas de calibração para a cromatografia de exclusão molecular em coluna HiPrep Sephacryl S300 16/60 HR	114
ANEXO IV - Proteínas envolvidas com o sistema de transferência de core conservado	115
ANEXO V - Alinhamento final mPAL_c1 e as sequências parentais: O padrão da sequência final em <i>scramble</i>	116
ANEXO VI - Alinhamento da mPAL_c1 e sequências de PAL/HAL com a estrutura resolvida	117
ANEXO VII - Análise de qualidade dos modelos de mPAL_c1, antes e após refinamento	120
ANEXO VIII - Deconvolução da curva de dicroísmo circular do agregado mPAL_c1 pelo software CDNN	121

Capítulo 1. REVISÃO TEÓRICA

Bioquímica, Mecanismos e Estrutura

A fenilalanina amônia liase (E.C. 4.3.1.24), PAL, foi inicialmente isolada de cevada (*Hordeum vulgare*) e caracterizada por Koukol e Conn (1961), tendo sido descrita com uma atividade de fenilalanina desaminase não-oxidativa de L-Phe em ácido trans-cinâmico e amônia. Em 1967, Moore e colaboradores demonstraram que tal enzima também estava presente em pelo menos 12 espécies de basidiomicetos.

Nas plantas superiores, a PAL se apresenta como o primeiro passo para a via dos fenilpropanóides, tendo um importante papel no metabolismo secundário envolvido com os processos de defesa (KOUKOL e CONN, 1961; WHETTEN e SEDEROFF, 1992). No entanto, na maioria dos microrganismos (fungos e bactérias), ela apresenta um papel catabólico na produção e reaproveitamento de fontes de nitrogênio e carbono (FRITZ *et al.*, 1976). A enzima se apresenta solúvel e citoplasmática na maioria dos casos.

Há um grande interesse nas pesquisas com a PAL, pois a sua atividade pode variar em magnitude de acordo com o tecido e estímulo (CAMM e TOWERS, 1973), tornando-a um elemento chave para a compreensão da fisiologia de diversas plantas e fungos, como também uma proteína marcadora para diversos processos. Além disso, é uma das poucas enzimas não-hidrolíticas com aplicações comerciais (AMBRUS *et al.*, 1978; EVANS *et al.*, 1987a; D'CUNHA *et al.*, 1994; MACDONALD e CUNHA, 2007).

A PAL purificada de diversas fontes apresenta uma massa molecular extremamente variável (Tabela 1), de 70 a 330 kDa (CAMM and TOWERS 1973; WATANABE *et al.* 1992; BRENDA-Enzyme, 2016). A enzima se apresenta em tetrâmeros (MACDONALD e CUNHA, 2007), sendo que dímeros tendem a formar um protômero com um único sítio ativo (CAMM e TOWERS 1973). Modificações pós-traducionais foram relatadas, como a glicosilação em *Triticum aestivum*, com 3 sítios confirmados para a formação de ligações N-glicosídicas (GOLDSON *et al.*, 2008); e a PAL isolada de batata que apresentou-se glicosilada com resíduos de manose em ligações N-glicosídicas nas cadeias laterais da enzima (SHAW *et al.* 1990). Por sua vez,

avaliando a enzima obtida de *Ustilago maydis*, Kim *et al.* (1996) não encontrou modificações, indicando que enzimas vegetais podem ser mais complexas que as fúngicas, devido às suas extensões características e potenciais modificações pós-traducionais.

Tabela 1. Principais características da fenilalanina amônia liase de diferentes fontes (Adaptado de BRENDA-enzyme.com: INFO E.C. 4.3.1.24).

ORGANISMO	K _M [mM]	MW [kDa]	pH ótimo	Inibidores	Localização
<i>Arabidopsis thaliana</i>	0,0428	ND	8,4-9,5	Ácido L-alfa-aminoóxi-beta-fenilpropiónico	ND
<i>Brassica juncea</i>	0,13	170	9,0	Zn ²⁺ , Cu ²⁺ , Ácido cinâmico, Ca ²⁺	Citosol
<i>Cucumis sativus</i>	0,043	316	8,8	Quercetina, D-Phe, Ácido cinâmico, ácido gálico, Campferol, o-cumarato	Microssoma
<i>Medicago sativa</i>	0,04	311	ND	Ácido sinápico, genisteína, ácido gálico, ferrulato, cumestrol, ácido cinâmico, ácido clorogênico, catequina, cafeato, 4-cumarato, daidzeína, formonetina, o-hidroxi-trans-cinamato, p-hidroxibenzoato	ND
<i>Nicotiana tabacum</i>	0,03	70-280	8,8	Tyr, 2-mercaptoetanol, o-cumarato, CN ⁻ , NaBH ₄	Citosol, Microssoma, Retículo endoplasmático
<i>Oryza sativa</i>	0,5	320	8,9	Quercetina, naringenina, ferrulato, cafeato, 4-cumarato, (3,4)-dihidroxibenzoato, ácido cinâmico	ND
<i>Petroselinum crispum</i>	0,0172	ND	10,0	N-metil-4-nitro-L-fenilalanina, ácido cinâmico, beta-(2-pirimidinil)-D,L-alanina, (N,N)-dimetil-4-nitro-L-fenilalanina	ND
<i>Rhizoctonia solani</i>	0,18	330	8,7-8,9	L-Cys, fenilpiruvato, DTT, dihidrocafeato, D-Phe, CN ⁻ , cafeato, 4-cumarato, 2-mercaptoetanol, GSH, NaBH ₄ , NaHSO ₃ , p-cloromercúriobenzoato	ND
<i>Solanum lycopersicum</i>	0,84	320	8,8	ND	Microssoma
<i>Solanum tuberosum</i>	0,038	330	ND	D-Phe, ácido cinâmico	Microssoma
<i>Ustilago maydis</i>	1,05	320	8,8-9,2	2-mercaptoetanol, 8-hidroxiquinolina, Ag ⁺ , ácido cinâmico, CN ⁻ , Cu ²⁺ , Hg ²⁺ , NaBH ₄ , p-cloromercúriobenzoato	ND
<i>Zea mays</i>	0,029	306	7,7	Ácido indol-butírico	Cloroplasto (tilacóide)

ND - Não determinado

A PAL possui uma grande diversidade ao longo dos táxons em que se apresenta, e pode estar presente sob a forma isozímica dentro de um mesmo organismo, sendo expressa em órgãos diferentes (CAMM e TOWERS, 1973; MACDONALD e CUNHA, 2004). Em *Arabidopsis thaliana* há 4 isoformas (AtPAL 1-4), com massas moleculares e afinidades pelo substrato (L-Phe) variáveis (COCHRANE *et al.*, 2004). Os valores de K_M para a PAL (Tabela 1) geralmente são relatados variando entre $1,1 \times 10^{-2}$ e 75,6 mM para seu substrato L-Phe (BRENDA-Enzyme, 2016). Esta enzima tende a não apresentar apenas um valor de K_M , mas este varia em virtude de vários fatores, como a concentração de produto, pH, conformação da enzima, e concentrações de substrato (CAMM e TOWERS, 1973).

A cinética enzimática da PAL parece ser anômala com um padrão bifásico nos gráficos de duplo-recíproco de Lineweaver–Burk, que são atribuídos à cooperativismo negativo (KHAN *et al.*, 1989). A sua cinética pode ser restituída à de Michaelis–Menten pela adição de D-Phe, um inibidor alostérico, que modifica as diferentes formas interconversíveis da enzima em apenas uma, padronizando a cinética final (HANSON e HAVIR, 1968; MACDONALD e CUNHA, 2007). Entre os microrganismos, a enzima extraída de *Rhodotorula* sp. parece seguir a cinética de Michaelis–Menten (WATANABE *et al.*, 1992).

A primeira estrutura tridimensional da PAL foi obtida por cristalografia de raios X e descrita por Calabrese *et al.* (2004), depositada sob o código 1T6P no banco de dados de PDB-RSC. A 1T6P, oriunda de *Rhodotorula glutinis*, é um homotetrâmero (Figura 1 A-B) que contém 716 resíduos por subunidade (76,88 kDa). Cada subunidade tem conformação de "cavalo marinho" e se arranja numa estrutura de cauda-cabeça entre as 2 subunidades, o que maximiza as interações, gerando uma área superficial final de contato igual a 58% da área total (MACDONALD e CUNHA, 2007).

A estrutura da PAL é similar ao *foldings* adotado pelas hisidina amônia-liases (HALs) (SCHWEDE *et al.*, 1999). A PAL possui ainda um núcleo central de hélices paralelas de diferentes comprimentos. A hélice mais longa da PAL espaça todo o monômero. Há apenas uma seção de folhas β maior que 3 resíduos; essa região forma um funil, que leva ao sítio ativo (Figura 1.B). A PAL difere da HAL por resíduos adicionais, principalmente concentrados nas regiões terminais, amino e carbóxi (CALABRESE *et al.*, 2004; PILBÁK *et al.*, 2006; MACDONALD e CUNHA, 2007),

Os mecanismos de reação da PAL são similares àqueles apresentados pela HAL (LANGER *et al.*, 2001; ASANO *et al.*, 2004). Inicialmente descobriu-se que havia por meio de modificações pós-traducionais a formação de uma dehidroalanina no sítio ativo da PAL (HANSON e HAVIR, 1968, 1970; HODGINS, 1971), e que este grupo prostético atuava com um eletrófilo, fazendo o intermédio da ligação da enzima com o nitrogênio quaternário da fenilalanina. No entanto, hoje sabe-se que este resíduo modificado é um intermediário de um grupamento prostético que atua como eletrófilo, o 3,5-dihidro-5-metildieno-4H-imidazol-4-ona (MIO) (ROTHER *et al.*, 2002; ALUNNI *et al.*, 2003), formado pela ciclização e desidratação do tripeptídeo interno Ala-Ser-Gly (Figura 3A). Dessa forma, as análises estruturais e experimentos posteriores confirmaram também que o grupamento eletrófilo MIO ataca o anel aromático do substrato numa reação do tipo Friedel–Crafts (SCHUSTER e RETEY, 1995; GLOGE *et al.*, 1998; LANGER *et al.*, 2001). A Figura 3B mostra que a ativação das ligações C-H ocorre e a clivagem traz ao mecanismo E1cb envolvido numa estabilização do carbânion intermediário (ALUNNI *et al.*, 2003).

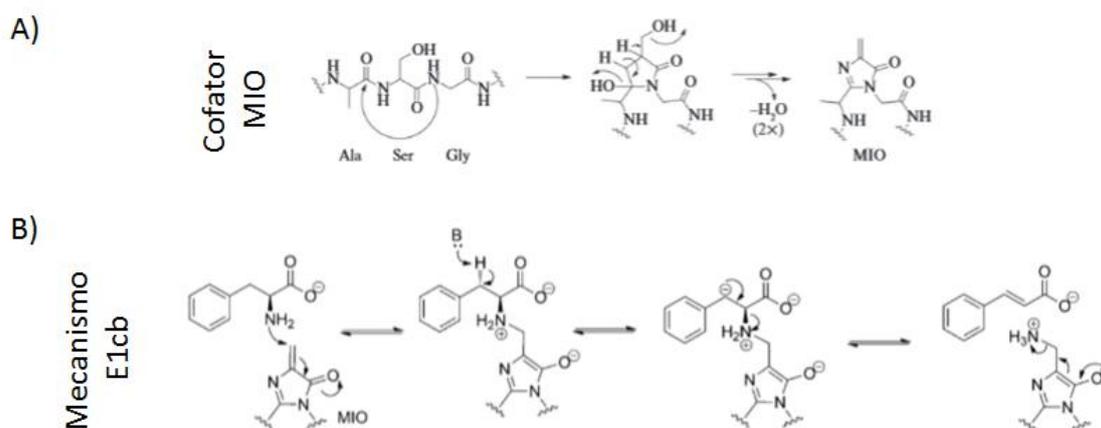


Figura 3. Mecanismo de reação da PAL. O cofator MIO é formado pela condensação do tripeptídeo ALA-SER-GLY (A) presente no sítio ativo e reage com a L-Phe, por meio da ligação inicial a seu nitrogênio quaternário, num mecanismo E1cb (B). Extraído de Lohman e Shen (2012).

A reação fisiológica reversa da PAL foi observada pela primeira vez por Hanson and Havir (1968) sob altas concentrações de NH_4^+ . Yamada *et al.* (1981) e Evans *et al.* (1987a, 1987b) chegaram a taxas de conversão de ácido trans-cinâmico e amônia em L-Phe próximas a 70%. O mecanismo da PAL parece ser do tipo Uni-Bi ordenado, com o ácido trans-cinâmico sendo liberado antes da amônia (WILLIAMS e HIROMS 1967).

Cianeto e NaBH_4 são inibidores não protéicos irreversíveis de PAL oriunda de batata, milho, tabaco, *Rhodotorula glutinis*, *Streptomyces verticillatus* (CAMM e TOWERS, 1973). Quando as PAL de batata e milho foram inativadas por NaB^3H_4 e logo em seguida hidrolisada, observou-se que o resíduo de alanina que compõe o cofator MIO da enzima foi marcado (HANSON e HAVIR, 1970), indicando que o cofator MIO é destruído por este tipo de compostos.

As PAL de cevada e *Rhodotorula* são sensíveis a reagentes sulfidrílicos (como o 2-mercaptoetanol, Tabela 1), enquanto que as preparações de batata, milho, *Ustilago*, e *Streptomyces* não o são (CAMM e TORRES, 1973). Há uma inibição universal da PAL por ácido trans-cinâmico, seu produto reacional, representando um mecanismo de *feedback* (MACDONALD e CUNHA, 2007). Compostos derivados do ácido cinâmico e compostos relacionados à sua estrutura, geralmente inibem a PAL (SATO *et al.*, 1982), Tabela 1. A PAL ainda pode ser inibida pelo álcool fenil etílico, sendo o fator chave para a inibição total da produção de pigmentos derivados do cinamato, como antocianinas (MO e SUNG, 2006).

Principais Vias

A PAL desempenha diferentes papéis em diferentes espécies (Figura 4). Nas plantas verdes, a PAL tem um papel chave na via dos compostos polifenólicos e está envolvido principalmente nos processos de defesa e proteção celular (anti-oxidação) (GHOLIZADEH e KOHNEHROUZ, 2010). Um dos principais processos em que a PAL é um agente chave, é o metabolismo da fenilalanina, que evidencia seu papel catabólico (ANEXO I). O processo catabólico da fenilalanina é diferente nos microrganismos e plantas, do que ocorre nos animais. Nessa via, a fenilalanina é convertida a ácido trans-cinâmico e amônia pela PAL ou PTAL. O ácido trans-cinâmico depois passa por pelo menos 4 reações até ser convertido a fumarato, que segue as

reações do ciclo do ácido cítrico. Numa via alternativa, o ácido trans-cinâmico pode ainda ser convertido a piruvato, através de 7 reações, e gera energia na forma de NADH e ATP, seguindo a via da glicólise. Finalmente, numa terceira via alternativa o trans-cinamato pode originar benzoato ou capsaicina (KEGG *Pathway* 00360). Estas vias objetivam geração de energia no organismo, através da degradação da fenilalanina absorvida ou do reciclo de peptídeos ao longo do ciclo celular.

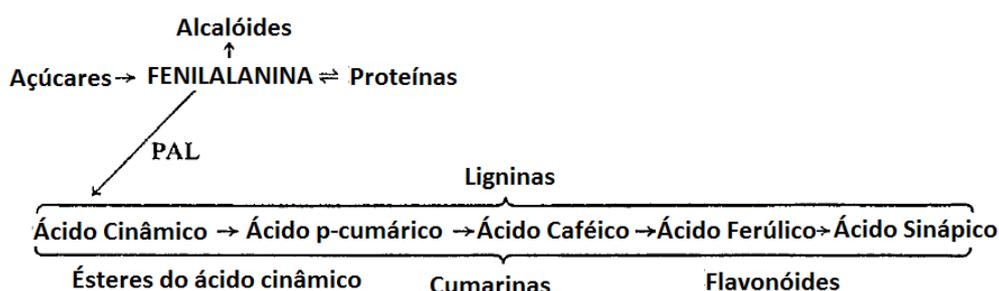


Figura 4. Principais compostos derivados da PAL (Adaptado de CAMM e TOWERS, 1973).

Tendo em vista a reação reversa que a PAL realiza, ainda pode-se ter a produção do aminoácido fenilalanina via condensação de trans-cinamato e amônia. Esse processo, além de útil para a célula é muito explorado industrialmente para produção de soluções enântio-puras de L-Phe (MACDONALD e CUNHA, 2007).

A PAL está envolvida também com a biossíntese dos fenilpropanóides (ANEXO II), que tem uma grande variedade de funções, desde estruturais à sinalização molecular (TAIZ e ZEIGER, 2006). Há ainda variações da via, onde o trans-cinamato, derivado da reação da PAL, pode ser direcionado como substrato para a coA ligase, e nesses casos forma-se o (E)-cinamoil-CoA que pode ser usado diretamente em diversas vias do metabolismo secundário *per si*, ou pode ser usado na produção de 4-cumaril-CoA por meio de uma hidroxilação e rumar para outras vias, tais como a dos curcuminóides (Figura 5). O (E)-cinamoil-CoA e seus compostos relacionados, como o 4-cumaril-CoA são precursores de numerosa classe de compostos, desde fenilpropanóides, flavonóides e benzenóides voláteis nas plantas, a policetídeos nas plantas e bactérias (MOORE *et al.*, 2002; KATSUYAMA *et al.*, 2009).

Nas plantas, o primeiro passo da maioria dessas vias, que é catalisado pela PAL, é a reação limitante. Uma vez que os níveis de expressão da PAL são fortemente regulados, e há ainda diversos processos de indução e repressão enzimática conhecidos (MACDONALD e CUNHA, 2007; LIU *et al.*, 2006).

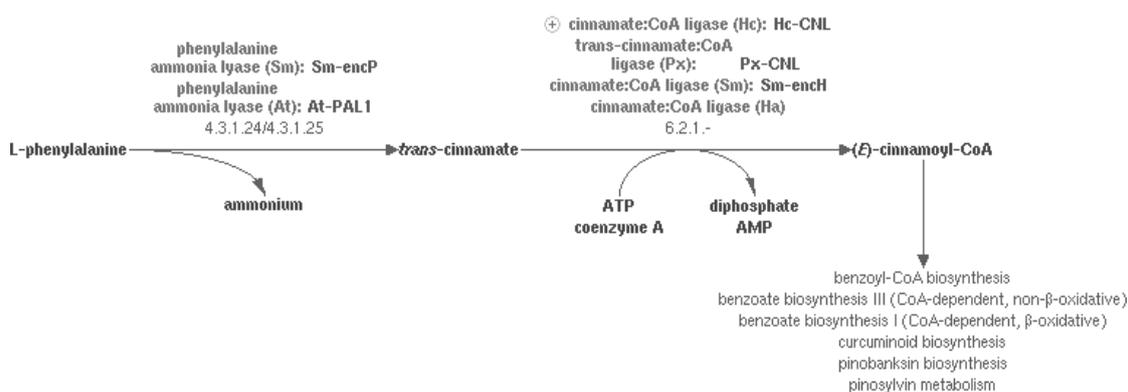


Figura 5. Via bioquímica do Cinamoil-coA (Extraído do site MetaCyc, entrada: PWY-6457).

Indução e Repressão: uma nova perspectiva sobre dados antigos

A PAL é uma enzima regulável (MARUSICH *et al.*, 1981; YAMADA *et al.*, 1981; NAKAMICHI *et al.*, 1983; EVANS *et al.*, 1987a, 1987b; ORNDORFF *et al.*, 1988), podendo ser reprimida ou induzida por diversos estímulos (Figura 6). Níveis aumentados da atividade de PAL são geralmente interpretados em termos de ativação enzimática, apesar da síntese *de novo* poder ser promovida (CAMM e TOWERS, 1973; FRITZ *et al.*, 1976; MACDONALD e CUNHA, 2007). As variações de atividade são devidas ao aumento de estabilidade do tetrâmero PAL, isso, pois pequenas proteínas poderiam se ligar ao mesmo que se encontra sob forma ativa, e convertê-lo numa forma inativa, pela desestabilização do tetrâmero nas regiões de extensão da enzima (Figura 1A). Aparentemente há ainda uma produção constitutiva desses inativadores de PAL e mudanças nos níveis da atividade de PAL são reflexo da mudança nas taxas de síntese deles e não necessariamente de enzima (CAMM e TOWERS, 1973).

Na figura 6 estão resumidos os principais mecanismos de estimulação da atividade da PAL, que podem ser resumidos em sua relação com o estresse oxidativo na

planta. Nossa hipótese, com base nas revisões de MacDonald e Cunha (2007) e de Camm e Towers (1973), é de que, com níveis aumentados de espécies reativas de oxigênio (ROS), o risco de: dano das membranas, deficiência de tradução e transcrição, alteração da atividade enzimática e indução de morte programada, pode induzir o sistema de produção de fenilpropanóides, que atuará como um tamponador de espécies oxidativas. Essa atividade anti-oxidante da PAL, é considerada não enzimática, e é desempenhada por compostos de alto peso molecular (como os polifenóis) que sequestram esses ROS e protegem as estruturas dos vegetais (OKADA e OKADA, 1998; GHOLIZADEH *et al.*, 2010).

Quando há maior irradiação solar na presença de dióxido de carbono, a taxa fotossintética se eleva. Isto induz um aumento da concentração de açúcares internamente, por meio da fotossíntese. Conseqüentemente, há uma maior taxa de metabolismo oxidativo, que gera maior quantidade de ROS e induz a ativação de PAL. Ela atua convertendo fenilalanina em trans-cinamato, que vai ser introduzido na via dos fenilpropanóides, que gera diversos compostos finais, como os polifenóis que podem atuar reduzindo o nível de ROS. O outro produto da PAL, a amônia, eleva o pH e reduz a taxa do metabolismo oxidativo.

Choques térmicos também induzem a atividade de PAL. Eles também estão correlacionados com maiores taxas de estresse oxidativo. Analogamente ao estímulo fotossintético, há uma indução de PAL para superprodução de polifenóis que reduzem os ROS, além da produção de amônia, que atua no equilíbrio do pH celular (ENGELSMAN 1969, 1970a, 1970b; CAMM e TOWERS, 1973).

Alguns fitohormônios alteram a atividade de PAL, sendo inibida por ácido indol-acético (IAA) e triptofano (INNERARITY *et al.*, 1972). Outros fitohormônios também atuam sobre a PAL, como: o 2,4D que reduz sua formação *in vivo* (DAVIES, 1972), as giberelinas que são capazes de promover um aumento na atividade de PAL, somente após exposição à luz (CHENG e MARSH, 1968), e abscisina 2 que aumenta os níveis de PAL mais rápido e em patamares maiores, mesmo no escuro (WALTON e SONDHEIMER, 1968). Apesar destes efeitos, pouco se sabe sobre o mecanismo de atuação destes fitohormônios sobre a PAL (MACDONALD e CUNHA, 2007; CAMM e TOWERS, 1973).

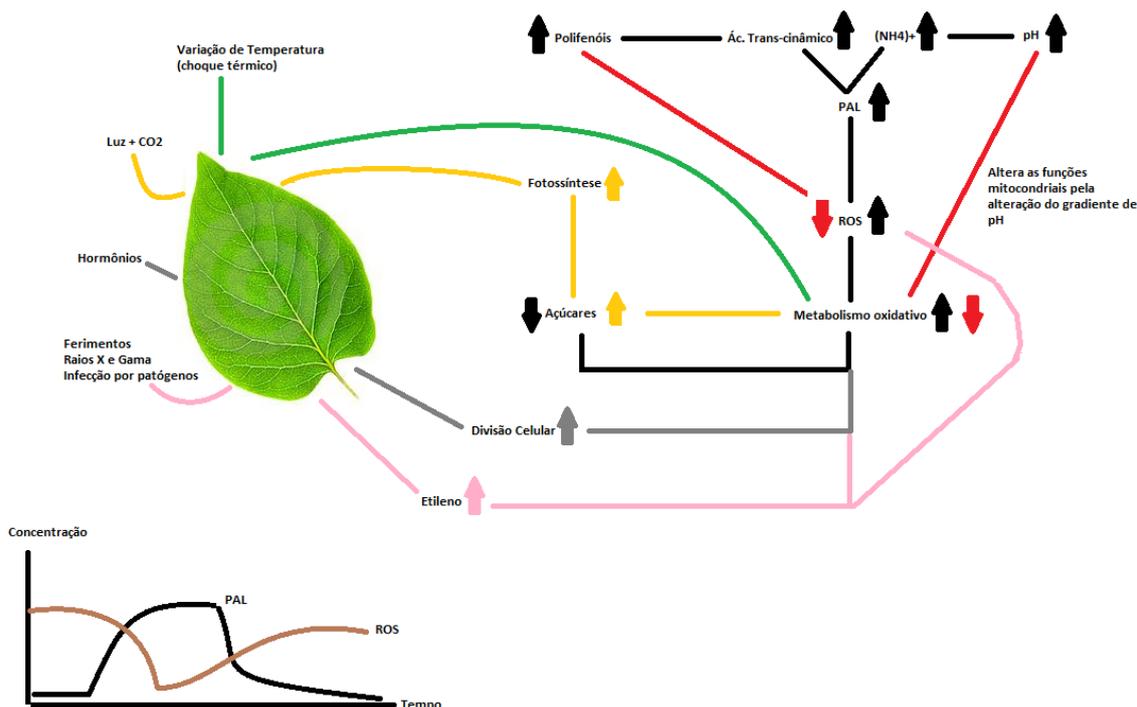


Figura 6. Vias de indução de repressão da atividade de PAL nas plantas. As vias foram construídas a partir da revisão dos mecanismos apresentados por CAMM e TOWERS (1973) e MACDONALD e CUNHA (2007). Vias separadas são mostradas em cores distintas, a via em preto tende a ser conservada para todos os estímulos. As setas indicam para enzimas, indução ou repressão, e para compostos, aumento ou redução da concentração. As espécies reativas de oxigênio foram abreviadas como ROS, e o potencial hidrogeniônico foi abreviado como pH. Há ainda um gráfico da atividade verificada de PAL e dos ROS *versus* o tempo, gerado por meio da interpretação de diversas referências da literatura.

Diversos estímulos (Raios X e Gama, ferimentos, infecção por patógenos) são capazes de elevar concentrações de etileno nas plantas (CAMM e TOWERS, 1973; HYUN *et al.*, 2011). A elevação desses níveis atua induzindo metabolismo oxidativo para combater os invasores, o que gera maior nível de ROS, que são tamponados com auxílio da PAL.

Apesar dos complexos mecanismos de indução e repressão que atuam sobre a PAL, seu papel metabólico ainda é bastante discutido. A atividade de PAL foi positivamente correlacionada com o teor de compostos polifenólicos e negativamente

correlacionada com níveis reduzidos de ácido clorogênico e cinamatos (CAMM e TOWERS, 1973). Altas atividades de PAL se correlacionam com a lignificação ativa de tecidos, apesar de tecidos não lignificados poderem apresentá-las, como no caso de aspargos e *Coleus* ou mesmo não apresentá-la como no caso de folhas de eucalipto e *Periploca graeca* (MACDONALD e CUNHA, 2007). Ao que parece, a PAL não atua somente na produção de flavonóides ou polifenóis, pois plantas com baixas atividades ou mesmo mutantes defectivos deste gene ainda apresentam tais compostos, além de cinamato, o que sugere a existência de vias alternativas, provavelmente dependentes de HAL ou enzimas como as aminoácido oxidases.

Diversidade e Evolução: estado da arte

A PAL, por ser a enzima chave da via dos fenilpropanóides, mostra-se muito interessante para análises filogenéticas. Essa enzima foi a primeira identificada na família dos genes de "defesa" das plantas (CRAMER *et al.*, 1989; EDWARDS *et al.*, 1985; KUHN *et al.*, 1984). Na maioria das plantas a PAL ocorre em famílias gênicas pequenas de 2 a 6 membros (CRAMER *et al.*, 1989; FRANK e VODKIN, 1991; GOWRI *et al.*, 1991; LEE *et al.*, 1992; LOIS *et al.*, 1989; MINAMI *et al.*, 1989; OHL *et al.*, 1990; SUBRAMANIAM *et al.*, 1993; TANAKA *et al.*, 1989; WANNER *et al.*, 1995; HAMBERGER *et al.*, 2007; BAGAL *et al.* 2011, 2012), havendo exceções notáveis em ambos os extremos, como a batata que contém mais de 40 genes codificantes de PAL (JOOS e HAHNBROCK, 1992) e o pinheiro que anteriormente acreditava-se conter apenas um único gene (WHETTEN e SEDEROFF, 1992), mas atualmente sabe-se que contém pelo menos 5 (BAGAL *et al.* 2011, 2012).

Em 1995, Wanner e colegas demonstraram que as sequências de aminoácidos das isoformas de PAL 1 e 2 presentes em *Arabidopsis thaliana* exibem identidade de 90% entre si, enquanto que com outras eudicotiledôneas e monocotiledôneas, gira em torno de 79% e 69%, respectivamente. A AtPAL3 é altamente variável, o que sugere uma origem ancestral anterior à divergência entre monocotiledôneas e dicotiledôneas. Dessa forma Wanner *et al.*(1995) sugerem que estudos utilizando táxons divergentes devam utilizar em suas análises sequências ortólogas às das AtPAL 1-2. Entretanto, pode ser possível que várias plantas contenham a ortólogas da isoforma AtPAL3, que

devido à sua divergência dos outros genes do grupo das angiospermas, ainda não foram identificados ou validados. Após uma análise da divergência evolutiva dessas isoformas (Tabela 2), por meio da substituição de aminoácidos entre duas sequências mostrou que a AtPAL1 é mais conservada e apresenta maior nível de similaridade dentre as plantas, o que reforça seu uso em trabalhos de filogenia utilizando este gene.

Tabela 2. Divergência evolutiva das isoformas de AtPAL1-3 e diferentes sequências de PAL. O número médio de substituições de aminoácidos por sítio entre 2 sequências está mostrado. A análise foi corrigida utilizando-se o modelo de Poisson proposto por Zuckerman e Pauling (1965). A taxa de variação dos sítios foi modelada com auxílio da distribuição Gama (parâmetro de forma = 5). A análise envolveu 38 sequências, onde todos os *gaps* foram eliminados. As análises foram conduzidas no software MEGA 6.06 (TAMURA *et al.*, 2013).

Espécie	AtPAL1	AtPAL2	AtPAL3	Acesso
<i>Aspergillus clavatus</i>	0,944	0,944	0,908	ACLA_080920
<i>Aspergillus fumigatus</i>	0,839	0,908	0,908	AFUA_2G09110
<i>Aspergillus niger</i>	0,873	0,944	0,873	ANI_1_738114
<i>Aspergillus oryzae</i>	0,714	0,743	0,657	AOR_1_934174
AtPAL1	0,000	0,112	0,225	AT2G37040; X62747
AtPAL2	0,112	0,000	0,208	L33678
AtPAL3	0,225	0,208	0,000	L33679
<i>Ipomoea batatas</i>	0,098	0,128	0,191	M29232
<i>Brachypodium distachyon</i>	0,208	0,208	0,260	100841454
<i>Cicer arietinum</i>	0,098	0,143	0,159	101496077
<i>Citrus clementina</i>	0,112	0,112	0,191	v10011134mg
<i>Citrus sinensis</i>	0,112	0,112	0,191	102620173
<i>Coprinopsis cinerea</i>	0,714	0,774	0,806	CC1G_06838
<i>Cucumis sativus</i>	0,069	0,128	0,191	101218144
<i>Fragaria vesca</i>	0,159	0,208	0,208	101315259
<i>Glycine max</i>	0,098	0,159	0,191	100811101
<i>M, sativa</i>	0,175	0,208	0,225	X58180
<i>Magnaporthe oryzae</i>	0,839	0,839	0,873	MGG_10036
<i>Medicago truncatula</i>	0,098	0,143	0,159	1g064090
<i>Neosartorya fischeri</i>	0,839	0,908	0,908	NFIA_084640
<i>Nostoc punctiforme</i>	0,774	0,839	0,714	R2068
<i>Oryza sativa japonica</i>	0,208	0,208	0,260	Os02t0627100-01
<i>Oryza sativa japonica</i>	0,225	0,225	0,278	4330037
<i>Populus trichocarpa</i>	0,054	0,112	0,191	0006s12870g
<i>Puccinia graminis</i>	0,714	0,743	0,714	PGTG_10006
<i>R, toruloides</i>	0,685	0,714	0,657	M18261
<i>Ricinus communis</i>	0,083	0,112	0,191	1355480
<i>Selaginella moellendorffii</i>	0,374	0,334	0,394	SELMODRAFT_450481
<i>Setaria italica</i>	0,225	0,225	0,278	101781518
<i>Solanum lycopersicum</i>	0,098	0,143	0,191	101243631

<i>Solanum tuberosum</i>	0,098	0,143	0,191	102596343
<i>Sorghum bicolor</i>	0,225	0,225	0,278	SORBI_04g026530
<i>Nicotiana tabacum</i>	0,098	0,143	0,191	M84466
<i>Uncinocarpus reesii</i>	0,982	0,908	0,873	UREG_04219
<i>Ustilago maydis</i>	0,944	0,944	0,982	UM00078.1
<i>Verticillium albo-atrum</i>	0,685	0,714	0,714	VDBG_08166
<i>Vitis vinifera</i>	0,128	0,128	0,175	100266593
<i>Zea mays</i>	0,242	0,242	0,296	100381820

Nas leveduras, como por exemplo, em *Rhodosporium* spp. (ANSON *et al.*, 1987) e *Rhodotorula rubra* (FILPULA *et al.*, 1988), a PAL é codificada como um único gene. Por sua vez, em fungos filamentosos a PAL é codificada por um número variável de cópias gênicas, podendo existir até quatro genes num mesmo genoma. Há relatos de genes de PAL, tanto de origem vegetal quanto fúngica, contendo íntrons (Tabela 3), sendo que os primeiros geralmente contém apenas 1 ou 2 íntrons, como no caso de *Arabidopsis thaliana* (WANNER *et al.*, 1995), enquanto os últimos tendem a ser poli-íntrônicos, principalmente no caso de leveduras, que possuem 5 íntrons (ANSON *et al.*, 1987) ou 6 íntrons (VASLET *et al.*, 1988). Já no caso das Pináceas e do fungo *Ustilago maydis* não foram encontrados íntrons (CAMPBELL, 1991; SEDEROFF *et al.*, 1994; KIM *et al.*, 2001).

Geralmente a PAL apresenta um *motif* característico: [GS]-[STG]-[LIVM]-[SAC]-S-G-[DH]-L-X-[PN]-L-[SA]-X_(2,3)-[SAGTVL] (HYUN *et al.*, 2011). Segundo Hyun e colegas (2011), os *motifs* da PAL se apresentam, nos fungos pelo menos, entre o 100° e 250° resíduos, exceto em clados específicos cuja organização tende a ser carboxi- ou amino-terminal, como respectivamente em *Neurospora* e *Neosartorya* (Tabela 3).

Tabela 3. Principais espécies de fungos e suas respectivas características da PAL

(Adaptado de HYUN *et al.*, 2011).

Espécie	Gene	Motif da PAL	Posição do motif	Proteína (res.)	Nº de íntrons
Ascomycota					
<i>Aspergillus clavatus</i>	ACLA_080920	GSISASGDLIPLSYIAA	123-139	664	1
<i>A. flavus</i>	AFL2G_05505	GSISASGDLIPLSYIAG	189-205	721	2
<i>A. flavus</i>	AFL2G_00533	GSISASGDLTPLAYIAG	202-218	714	1
<i>A. flavus</i>	AFL2G_06214	GSISASGDLSPYIYISG	155-171	671	3
<i>A. fumigatus</i>	Afu2g09110	GSISASGDLMPLSYIAG	181-197	728	2
<i>A. nidulans</i>	ANID_03897	GSISASGDLTPLAYIAG	182-198	687	1
<i>A. nidulans</i>	ANID_06075	GSISASGDLMPLSYIAG	187-203	702	2

<i>A. niger</i>	e_gw1_15.39	GSISASGDLTPLAYIAG	194-210	719	1
<i>A. niger</i>	e_gw1_3.237	GSISASGDLSPSYIAG	181-197	720	2
<i>A. niger</i>	fge1_pm_C_13000016	GTISASGDLMPLAYVVG	184-200	704	2
<i>A. oryzae</i>	AO090005000532	GSISASGDLTPLAYIAG	202-218	714	1
<i>A. oryzae</i>	AO090011000788	GSISASGDLIPLSYIAG	189-205	721	2
<i>A. oryzae</i>	AO090026000586	GTISASGDLMPLAYVTG	183-199	696	2
<i>A. oryzae</i>	AO090701000601	GSISASGDLSPYIAG	228-244	744	4
<i>Fusarium. graminearum</i>	FGSG_09311	GTISASGDLMPMSYIAG	180-196	721	0
<i>F. oxysporium</i>	FOXG_05927	GTISASGDLMPLSYIAG	180-196	750	0
<i>F. verticillioides</i>	FVEG_03798	GTISASGDLMPLSYIAG	180-196	724	0
<i>F. verticillioides</i>	FVEG_10552	GSISASGDLIPLSYIAG	196-212	706	1
<i>Magnaporthe oryzae</i>	MGG_10036.7	GSISASGDLALAWIAA	207-223	627	0
<i>M. poae</i>	MAPG_07598.1	GSISASGDLSPSYIAG	198-214	730	2
<i>Neurospora crassa</i>	NCU09391.5	GSISASGDLSTLSYIAG	610-623	763	1
<i>Neosartorya fischeri</i>	NFIA_084640	GSISASGDLMPLSYIAG	48-64	595	0
<i>Uncinocarpus reesii</i>	UREG_04219.1	GTISASGDLMPLAYIVG	185-201	710	2
<i>Verticillium albo-atrum</i>	VDBG_08166.1	GTISASGDLMPLSYIAG	181-197	676	1
Basidiomycota					
<i>Laccaria bicolor</i>	LACBIDRAFT_291120	GTISASGDLAPLSYIAG	163-179	688	5
<i>L. bicolor</i>	LACBIDRAFT_184628	GSISASGDLSPSYIAG	201-217	731	11
<i>Puccinia triticina</i>	PTTG_02413.1	GSISASGDLMPLSYVAA	175-191	653	13
<i>P. graminis tritici</i>	PGTG_12283.2	GSISASGDLMPLSYVAA	190-206	691	13
<i>Rhodosporidium toruloides</i>	AAA33883	GTISASGDLSPSYIAA	207-213	693	6
<i>R. toruloides</i>	P11544	GTISASGDLSPSYIAA	184-200	716	6
<i>Rhodotorula graminis</i>	CAD23828	GSISASGDLSPSYIAG	213-229	713	1
<i>R. mucilaginosa</i>	CAA31486	GTISASGDLSPSYIAA	213-229	720	5
<i>Ustilago maydis</i>	UM00078	SSISASGDLSPSYVAG	201-217	724	0

Não existem estudos sobre a relação filogenética das PALs vegetais e fúngicas, que sejam definitivos. A PAL parece possuir uma história evolutiva bastante complexa e altamente dependente de transferências horizontais (Figuras 7 e 8). Emiliani e colaboradores (2009) reportam que num cenário mais parcimonioso a PAL surgiu primitivamente em organismos bacterianos do solo (provavelmente uma *Nostocale*) que transferiram o gene para um ancestral dos fungos do grupo *Dikarya* que posteriormente transferiu esse gene a um ancestral das plantas terrestres, como mostrado em detalhes na Figura 7. Tal argumento propõe de modo interessante, que plantas primitivas necessitariam de adaptações metabólicas para proteção contra estresses abióticos, tais como: UV, dessecação e outros, facilmente contornáveis pela via dos fenilpropanóides. Além disso, parece importante destacar que o gene codificante de HAL parece ter sido perdido muito cedo na evolução do grupo das plantas, dessa forma, a via dos fenilpropanóides não poderia ter emergido sem a transferência horizontal de genes codificadores de PAL.

Hyun *et al.* (2011) observaram que nos fungos a PAL pode ser dividida em 3 grupos principais: Ascomycota I, II e Basidiomycota. O grupo Ascomycota I é mais intrinsecamente relacionado aos Basidiomycota, que aos Ascomycota II. A grande

diversidade da sequência de PAL dentro dos fungos é mostrada no cladograma (Figura 8) e revela uma divergência proporcionalmente grande (variando entre 33 e 77% de similaridade) dentro dos diferentes grupos. A maior identidade verificada dentro dos Ascomycota foi de 65%, enquanto que para os Basidiomycota foi de 97%; cruzando-se os grupos o maior nível de identidade alcançado foi de 41%.

Nos fungos, diferentemente das plantas, a PAL tem papel catabólico na obtenção de carbono e nitrogênio de fontes de aminoácidos externas, entretanto, outras enzimas podem realizar a mesma função neste grupo de organismos, como a Fenilalanina aminotransferase ou a Aminoácido oxidase. Sendo assim, Hyun e colaboradores (2011) ainda questionam qual seria o papel evolutivo da PAL para estes organismos e qual sua vantagem evolutiva em relação aos demais, chegando a conclusão de que aparentemente há alguma vantagem evolutiva na produção de ácido trans-cinâmico que pode ser importante no ciclo de vida de diversos fungos. Além disso, Emiliani *et al.* (2009) argumentam que os produtos da PAL poderiam atuar na produção de compostos antimicrobianos.

A expansão e diversificação funcional das famílias gênicas de PAL parece ser crítica na evolução das plantas e na sua habilidade de se adaptar aos estresses biótico e abiótico (HAMBERGER *et al.*, 2007; EMILIANI *et al.*, 2009). *Populus* e *Arabidopsis* são gêneros relativamente relacionados ao grande clado Eurosídios do grupo das Eudicotiledôneas, e compartilham muitas estratégias de desenvolvimento e defesa, ainda que possuam histórias de vida e adaptação fortemente contrastantes. Numa análise superficial dos grandes grupos gênicos relacionados à via dos fenilpropanóides, Hamberger *et al.* (2007) determinaram que para a PAL há uma clara separação entre as monocotiledôneas e dicotiledôneas (Figura 8). Assim, a diversificação dos genes da PAL aparentemente ocorreu após o desdobramento de monocotiledôneas e eudicotiledôneas a partir de um único progenitor comum. A presença de ambos, *Populus* e *Arabidopsis*, num grupo estreitamente relacionado, mas com divergências óbvias, ainda sugere que a emergência destes dois clados foi um evento precoce na evolução das eudicotiledôneas.

Enquanto isso, Wanner e colaboradores (1995) especulam que as sequências de AtPAL podem ter tido origens diversas e que a AtPAL3 pode ter tido uma origem diferente das AtPAL1-2, datando de um período anterior à divergência entre monocotiledôneas e dicotiledôneas (Tabela 2). Bagal e colaboradores (2012) ainda identificaram que os genes de PAL de gimnospermas mostram alta similaridade com

genes de táxons não-gimnospérmicos, o que pode refletir-se na análise filogenética (Figura 8). Os sítios relativos à ligação ao substrato e atividade catalítica são aparentemente altamente conservados, juntamente com resíduos identificados como "marcadores" para PAL, previamente descritos por Xu *et al.* (2008) e Calabrese *et al.* (2004).

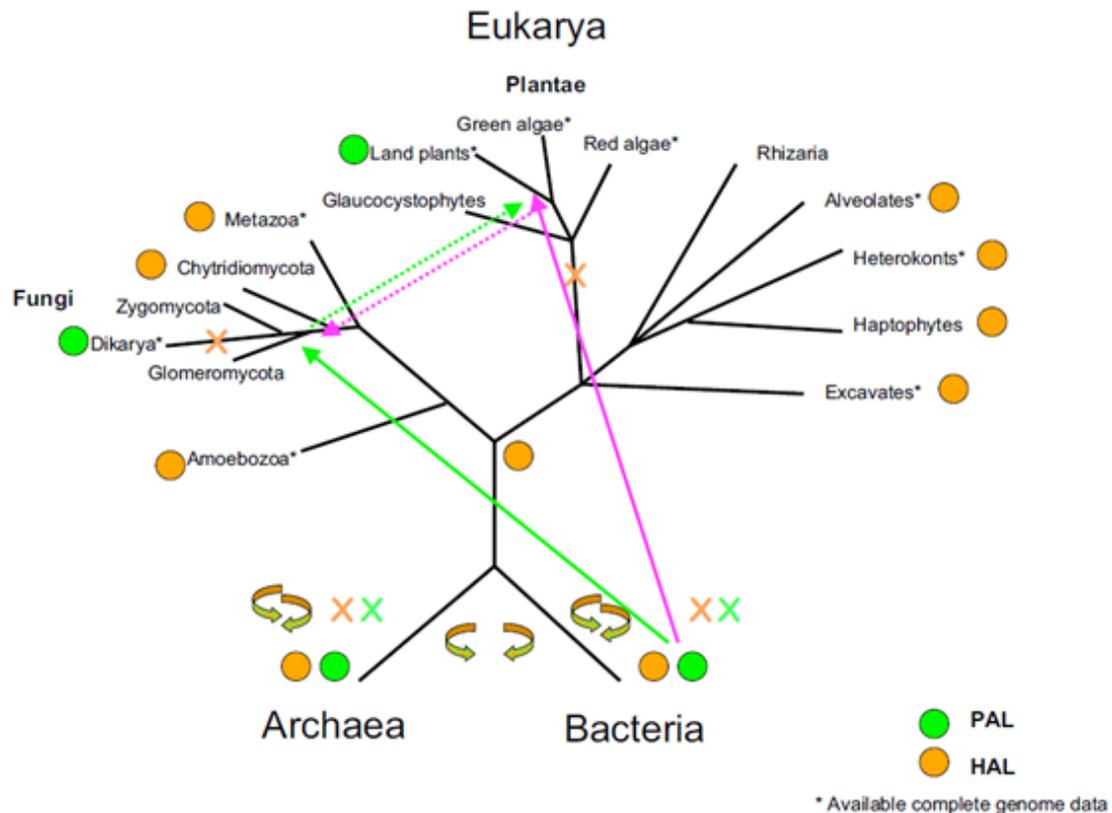


Figura 7. Cenário evolutivo para surgimento da PAL vegetal e fúngica. As cruces mostram possíveis deleções gênicas, enquanto as setas mostram possíveis caminhos para transferências gênicas ao longo das histórias evolutivas desses organismos. Dentro dos grupos bacterianos e das arqueas o fluxo gênico não pode ser determinado facilmente e por isso foi mostrado como um sistema de setas circulares. Ainda há a presença e ausência dos genes PAL e HAL, exibido pela presença de círculos com as cores previamente legendadas. (Adaptado de EMILIANI *et al.*, 2009).

Bagal *et al.* (2012) discutem que um clado de Gimnospermas 1 separou-se dos outros grupos vegetais com um ancestral anterior, enquanto que as sequências restantes foram separadas em 2 grupos (Angiospermas e Gimnospermas 2 e 3). Sendo que o

Um mapeamento por parcimônia, realizado por Bagal e colaboradores (2011, 2012) sugere origens sucessivas dos três clados de gimnospermas, no que se refere ao gene PAL, anteriores à origem do clado das angiospermas. Plantas ancestrais com sementes tinham três genes distintos para PAL que foram conservados em gimnospermas, mas dois destes genes ancestrais foram perdidos na linhagem das angiospermas após a divergência. Além disso, os genes PAL também sofreram diversificação, mais recentemente, no clado dos pinheiros. Há ainda uma tendência a perda de diversidade em angiospermas (OKADA *et al.*, 2008). Eventos de divergência recente agrupam genes que mostram propriedades bioquímicas, moleculares e catalíticas distintas (KUMAR e ELLIS, 2001).

Apesar de todos os estudos até então conduzidos, não houve uma avaliação completa da filogenia da PAL, no que tange os clados divergentes de diferentes reinos, como fungos e plantas. Ainda há uma grande divergência metodológica que atualmente, devido à grande distância temporal destes estudos, está em defasagem. Os estudos que embasam toda a teoria de filogenia que concernente à PAL foram centrados em árvores feitas com a metodologia de agrupamento de *Neighbor-Joining*, que foi substituída e provada ser inferior às de *Maximum likelihood* (ML) e Bayesiana (YANG, 2007; NINIO *et al.*, 2006; HOLDER e LEWIS, 2003). Outros pontos que enfraquecem o uso destes dados são o baixo nível de confiança do suporte dos nós das árvores, que em sua maioria foram feitos com uma análise de *bootstrap* que utiliza apenas 100 pseudorreplacas, 10 vezes menor que padrão ouro da técnica (YANG, 2007). Nenhum dos estudos observados até então, levou em conta as características que evolução direcionada de determinados grupos tende a apresentar, e por consequência ainda não houve determinação de modelos de substituição e correção específicos para os alinhamentos desenhados, quando utilizou-se modelos para correção da topologia das árvores (BAGAL *et al.* 2011, 2012) houve um erro ao se desconsiderar a bondade de ajuste do modelo aos dados, uma vez que não houve os testes dos critérios informativos de Akaike (AKAIKE, 1974; POSADA e BUCKLEY, 2004) e Bayesiano (CHEN e GOPALAKRISHNAN, 1998; POSADA e BUCKLEY, 2004; STAFYLAKIS *et al.*, 2009). Desta forma, estudos conduzidos com os padrões do estado da arte são necessários.

Aplicações biotecnológicas

A PAL é uma enzima interessante do ponto de vista biotecnológico, tanto por sua reação direta quanto inversa, e é uma das poucas enzimas não hidrolíticas sendo atualmente produzida e utilizada pela indústria para diversas aplicações, como síntese e diagnóstico. Apesar de sua ampla distribuição dentre os vegetais e clados microbianos, a fonte mais utilizada ainda tem sido as enzimas provenientes do gênero *Rhodotorula* sp. (MACDONALD e CUNHA, 2007). Em Janeiro de 2016, o custo de 2 U de PAL obtida do fungo *R. glutinis* na empresa Sigma-Aldrich (código do produto P1016), uma das únicas fornecedoras dessa enzima no país, era de R\$ 906,00.

Além de sua aplicação prática na determinação dos teores de L-Phe em plasma, soro, urina e outros fluidos corporais (WATANABE *et al.*, 1992); Barrat *et al.* (1976) utilizaram-na para alterar as concentrações desses aminoácidos no plasma de macaco-esquilo. A fenilcetonúria é uma doença cuja inatividade congênita da fenilalanina-4-hidroxilase, leva ao acúmulo de L-Phe no organismo (SARKISSIAN e GAMEZ, 2005), o que desencadeia sintomas de reação neurotóxica (SCRIVER e KAUFMAN, 2001; DONLON *et al.*, 2004). Uma de suas vertentes terapêuticas mais modernas utiliza uma forma recombinante da PAL (SARKISSIAN e GAMEZ, 2005), para converter o excesso de L-Phe em metabolitos inofensivos, t-CA e amônia, que são excretados na urina. A administração da enzima inicialmente foi proposta sob a forma de uma injeção subcutânea de PAL (LÓPEZ-VILLALOBOS *et al.*, 2014). A forma recombinante da PAL, quando administrada oralmente tem efeito metabólico reduzido ao longo de múltiplas aplicações, provavelmente devido a uma resposta imune e proteólise extensiva no lúmen intestinal, no entanto, estudos viabilizando sua forma peguilada tem sido conduzidos com sucesso (GÁMEZ *et al.*, 2005).

A PAL na indústria é utilizada na produção de L-Phe a partir de t-CA pela sua reação inversa (YAMADA *et al.*, 1981; HAMILTON *et al.*, 1985; EVANS *et al.*, 1987a, 1987b). A L-Phe tem sido utilizada como um precursor de aspartame, e o processo utilizando a PAL como reação inicial foi desenvolvido pela *Genex Corporation* nos Estados Unidos (FINKELMAN e YANG, 1985), eliminando o passo de esterificação química de L-Phe com metanol na síntese de aspartame.

Na engenharia de alimentos, a inibição da atividade da PAL tem sido correlacionada com o aumento da vida de prateleira e qualidade visual global de vários alimentos, como o alface processado (LÓPEZ-GALVEZ *et al.*, 1996) e aspargos frescos recém cortados, pós tratamento com ozônio aquoso (AN *et al.*, 2007).

Ainda existem aplicações industriais potenciais, como reações em fluxo com meio orgânico, que derivam de sua robustez. Diferentemente da maioria dos estudos sobre enzimas em meios não aquosos, que envolvem estruturas monoméricas catalisadoras de reações hidrolíticas (BATTISTEL e BIANCHI, 1994; ZAKS e KLIBANOV, 1984; MOZHAEV *et al.*, 1991), a PAL se mostrou ativa, como uma proteína oligomérica, e estável em solventes orgânicos, como o n-octanol minimamente hidratado (98%) (GARETH REES e JONES, 1996, 1997), sendo ainda capaz de reagir com éster metílico de trans-cinamilo (t-CM), formando éster metílico de L-Phe (D'CUNHA *et al.*, 1994, 1996).

Outras aplicações da PAL ainda podem ser citadas, como: a preparação de dietas com baixo conteúdo de L-Phe (HSIA e HOLTZMANN, 1973), tratamento de certos tumores neoplásicos - adenocarcinoma BW10232 e melanoma B16 (FRITZ *et al.*, 1976) e tratamento para leucemia (ABELL *et al.*, 1972).

A biologia sintética e design de proteínas

A biologia sintética consiste na aplicação de conceitos da engenharia na biologia e visa não só alterar sistemas biológicos naturais, mas construir racionalmente sistemas complexos de componentes bem compreendidos (MacDONALD *et al.*, 2011; PLEISS, 2011). Parte dessa ciência é a produção racional de enzimas com novas funções catalíticas (JIANG *et al.*, 2008; ROTH LISBERGER *et al.*, 2008; SIEGEL *et al.*, 2010) mesmo que possuam atividades muito mais baixas do que as enzimas que ocorrem naturalmente.

A concepção computacional de proteínas é geralmente dividida em duas etapas (BUTTERFOSS e KUHLMAN, 2006), numa primeira etapa é necessário encontrar a base sobre a qual o *design* será produzido, que pode ser gerada artificialmente ou derivada de estruturas de proteína resolvidas experimentalmente; na segunda etapa é

necessário encontrar sequências de aminoácidos capazes de se dobrar na estrutura necessária, minimizando a energia livre.

Entre os métodos mais populares do desenho dinâmico de proteínas está a adoção do "*design* positivo", onde encontra-se os rotâmeros de cadeia lateral que minimizam a energia potencial da estrutura receptora por meio de métodos de anelamentos ou simulações (ROHL *et al.*, 2004; DAHIYAT e MAYO, 1996). Outra abordagem é a concepção de enzimas de modo reverso, para tal é necessário determinar os grupos funcionais químicos que são necessários para estabilização do estado de transição, e em seguida, encontrar as cadeias laterais compatíveis testando suas conformações, e, finalmente, encontrar uma estrutura receptora compatível (ZANGHELLINI *et al.*, 2006). Complementando esse método há a determinação das possíveis interações moleculares capazes de estabilizar o estado de transição por meio da comparação com estruturas de enzimas conhecidas.

Apesar de todos os métodos disponíveis, métodos simples e acessíveis à pesquisadores com pouca experiência bioinformática ou sem acesso à servidores para o processamento de grande volume de dados são ainda necessários.

Objetivos e Justificativas

O presente estudo justifica-se pela produção de uma nova enzima por meio de um novo protocolo com potencial de auxiliar a engenharia genética reversa.

Tendo em vista que a PAL é homóloga da HAL, que está envolvida no catabolismo de histidina e é comum em procariotas e eucariotas (ROTHER *et al.*, 2001; TAYLOR *et al.*, 1990); e que há um consenso de que a PAL original se desenvolveu a partir de uma HAL ancestral, quando fungos e plantas divergiram dos outros reinos (FERRER *et al.*, 2008). Este estudo é uma prova de conceito da hipótese de que: enzimas relacionadas podem ser modificadas ou fusionadas, por meio de *design in silico*, para obtenção de enzimas com atividades específicas de interesse.

Assim sendo, o objetivo geral do trabalho foi a obtenção de uma enzima utilizando biologia sintética, por meio da inserção de um *core* catalítico conservado em detrimento de posições variáveis, cuja sequência foi obtida a partir de um *data set* de

enzimas PAL, num arcabouço de uma HAL filogeneticamente distante que por sua relação ancestral permitiria uma "compatibilidade". Dessa forma, objetivou-se deslocar a atividade da enzima original (HAL) tornando-a ativa em presença de L-Phe e reduzindo sua afinidade pelo substrato original, L-His.

Como objetivos específicos tivemos:

- compreensão e análise filogenética completa, por meio de abordagens de evolução molecular, de um *data set* de sequências ortólogas de PAL;
- desenho *in silico* de uma enzima sintética, determinando um protocolo específico;
- síntese comercial do gene;
- clonagem do gene e expressão heteróloga em hospedeiro bacteriano;
- purificação enzimática;
- caracterização da atividade enzimática com relação aos substratos L-Phe e L-His;
- realização de testes de qualidade da enzima final, quanto à sua estrutura;
- comparação da atividade da enzima sintética com a enzima comercial de *Rhodotorula glutinis*;
- análise do efeito do íon amônio.

Capítulo 2. *INSIGHTS* SOBRE A EVOLUÇÃO DA FENILALANINA AMÔNIA LIASE (PAL): EVENTOS DE RECOMBINAÇÃO PODERIAM TRAZER DIVERSIDADE À SUA HISTÓRIA?

MATERIAL E MÉTODOS.

Utilizou-se trinta e duas sequências de proteínas PAL ortólogas, baixadas do KEGG - Enciclopédia de Genes e Genomas de Kyoto (<http://www.kegg.jp/>), disponíveis na Tabela 4. A referência filogenética utilizada foi a estrutura filogenética obtida a partir de uma abordagem conjunta utilizando os bancos de dados: KEGG e do *Tree of Life Web Project* (MADDISON *et al.*, 2007), que foi usado para comparações. A proteína actina foi utilizada como âncora filogenética, uma vez que é conservada e evolui lentamente (COX *et al.*, 1995; HELGASON *et al.*, 2003; SIN *et al.*, 2007; ZHANG *et al.*, 2010; LAHR *et al.*, 2011), na comparação das taxas de evolução e alguns outros parâmetros da PAL. Utilizou-se a sequência de PAL oriunda de *Nostoc punctiforme* como um grupo externo para as nossas análises. Os grupos de sequências foram segregados através de distinção filogenética previamente determinada, excluindo os fungos, que foram classificados quanto ao seu papel ecológico: endófitos (compreendendo aqueles do gênero *Aspergillus*) (ARAUJO *et al.*, 2010), de vida livre (*Uncinocarpus reesii* e *Coprinopsis cinerea*) (SHARPTON *et al.*, 2009; STAJICH *et al.*, 2010) e patógenos (*Ustilago maydis*, *Magnaporthe oryzae*, *Puccinia graminis*, *Verticillium albo-atrum*) (DEAN *et al.*, 2005; KÄMPER *et al.*, 2006; KLOSTERMAN *et al.*, 2011). Todos os experimentos foram realizados *in silico* e seguiram os parâmetros pré-estabelecidos em outros estudos (COSCOLLÁ *et al.*, 2006; COSTA *et al.*, 2010; COSTA *et al.*, 2012). As análises evolutivas filogenéticas e moleculares foram realizados utilizando o software MEGA versão 6.03 (TAMURA *et al.*, 2013).

As sequências de aminoácidos foram alinhadas pelo algoritmo *Clustal Ω* (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), utilizando valores de penalidade estabelecidos para *-gap opening* e *-gap extension* de 15,0 e 6,6, respectivamente. Obteve-se os blocos de resíduos contíguos conservados, com elevada estringência a partir da curação de G-blocos no servidor PhyML 3.0 (GUINDON e GASCUEL 2003), utilizando-se os parâmetros mínimos pré-estabelecidos pelo próprio servidor.

O modelo mais apropriado de substituição de aminoácidos e de pontuações de probabilidade foi avaliado com 2 diferentes softwares: Topali V2.5 (MILNE *et al.*, 2009) e JModel Test (POSADA, 2008). A qualidade do ajuste do melhor modelo de substituição foi determinada por meio da análise dos valores dos critérios informativos de Akaike (AICc) (AKAIKE, 1974; POSADA e BUCKLEY, 2004) e Bayesiano (CHEN e GOPALAKRISHNAN, 1998; POSADA e BUCKLEY 2004; STAFYLAKIS *et al.*, 2009), sendo que o modelo que rendeu os melhores valores em ambos os critérios foi estabelecido como o mais acurado.

A reconstrução filogenética foi realizada utilizando o método de Máxima Verossimilhança (ML). A topologia foi testada pelo método de *Bootstrap* com 1000 pseudo-replicatas. A matriz de substituição mais precisa para as sequências de aminoácidos foi o modelo LG (LE e GASCUEL, 2008) com adição de frequências discretas (+F), Gama distribuído com 5 categorias distintas contendo a correção para sítios invariantes (+G+I). As lacunas e dados faltantes foram completamente eliminados das análises finais. A inferência da árvore foi feita através do método heurístico ML, baseado no algoritmo de *Nearest-Neighbor-Interchange* (NNI). A árvore inicial foi feita automaticamente com o algoritmo de *Neighbor-Joining* (NJ) / Bio(NJ). Foi utilizado um filtro de *swaps* no ramo com poder de discriminação máximo.

As matrizes de distância foram fornecidas pela análise de divergência evolutiva em rede (*Net Evolutionary Divergence*) entre grupos de sequências usando o modelo de correção de *Poisson* (ZUCKERKANDL e PAULING, 1965). A variação entre os locais foi modelada com uma distribuição gama (parâmetro de forma = 5).

Os eventos de reticulação foram evidenciados através de uma análise de *Neighbor-net* (BRYANT e MOULTON, 2004) e convertidos em um gráfico de *Splits* com auxílio do software SplitsTree-versão 4.10 (HUSON e BRYANT, 2006). O método de *Neighbor-net* foi baseado nas matrizes de pares de distâncias das sequências completas alinhadas da PAL com exclusão de lacunas e sítios constantes ou não informativos para parcimônia; o modelo de distribuição de *Poisson* foi utilizado para o cálculo de correção das matrizes de distância (ZUCKERKANDL e PAULING, 1965).

Tabela 4. Sequências ortólogas de PAL e Actina utilizadas neste estudo. As principais características da proteína PAL são mostradas, incluindo a proporção de aminoácidos carregados negativamente (D+E) e positivamente (R+K). O comprimento e o peso molecular (MW) são dadas em resíduos e Daltons (Da), respectivamente. Os grupos de vida livre (Free life), patógenos e endófitos referem-se a subgrupos de fungos, bem como os grupos Monocot e Dicot referem-se respectivamente a plantas monocotiledôneas e dicotiledôneas.

Código	Grupo	Espécie	Código de Acesso no KEGG		Características Bioquímicas da PAL								
			Actina	PAL	NRes. (aa)	pI	MW	ININ	AI	GRAVY	D+E	R+K	(-)/(+)
osa	Monocot	<i>Oryza sativa japonica</i>	Os05g0106600	4330037	713	5,84	76581,4	33,17	91,91	-0,078	81	71	1,14
dosa	Monocot	<i>Oryza sativa japonica</i>	Os01g0964133	Os02t0627100-01	713	5,9	76980,9	33,58	90,93	-0,126	85	74	1,16
sbi	Monocot	<i>Sorghum bicolor</i>	Sb09g000750	SORBI_04g026530	714	6,05	76823,9	33,52	92,44	-0,08	81	73	1,11
sita	Monocot	<i>Setaria italica</i>	101786149	101781518	713	5,89	76778,8	33,71	93,38	-0,08	82	72	1,14
zma	Monocot	<i>Zea mays</i>	100191561	100381820	715	5,89	76932,9	34,43	92,31	-0,083	82	72	1,14
bdi	Monocot	<i>Brachypodium distachyon</i>	100831731	100841454	712	6,26	76629,6	34,01	90,63	-0,103	81	75	1,08
rcu	Dicot	<i>Ricinus communis</i>	503940	1355480	714	6,18	77719,8	34,62	88,45	-0,171	81	73	1,11
pop	Dicot	<i>Populus trichocarpa</i>	0001s31700g	0006s12870g	714	6,1	77715,5	32,98	89,1	-0,202	82	73	1,12
gmx	Dicot	<i>Glycine max</i>	100789000	100811101	712	6,07	77794,6	32	90,44	-0,201	80	70	1,14
cam	Dicot	<i>Cicer arietinum</i>	101503086	101496077	723	6,13	78785,8	32,61	92,16	-0,182	78	69	1,13
mtr	Dicot	<i>Medicago truncatula</i>	3g095530	1g064090	712	5,93	77372,1	34,48	91,81	-0,162	80	69	1,16
csv	Dicot	<i>Cucumis sativus</i>	101226611	101218144	712	5,8	77793,6	35,29	92,5	-0,194	84	71	1,18
ath	Dicot	<i>Arabidopsis thaliana</i>	AT5G09810	AT2G37040	725	5,9	78725,7	31,37	90,03	-0,156	82	70	1,17
cic	Dicot	<i>Citrus clementina</i>	v10028665mg	v10011134mg	757	6,06	82467,8	30,42	89,46	-0,198	85	74	1,15
cit	Dicot	<i>Citrus sinensis</i>	102614385	102620173	721	6,16	78472,2	28,95	89,46	-0,233	82	73	1,12
vvi	Dicot	<i>Vitis vinifera</i>	100246825	100266593	710	6,17	77079,9	33,38	89,7	-0,17	82	73	1,12
sly	Dicot	<i>Solanum lycopersicum</i>	101260631	101243631	721	6,29	78355,5	33,68	91,19	-0,18	80	74	1,08
sot	Dicot	<i>Solanum tuberosum</i>	102584969	102596343	722	6,04	78488,7	33,77	90,93	-0,155	81	72	1,12

fve	Dicot	<i>Fragaria vesca</i>	101292078	101315259	724	6,1	78985	34,14	90,95	-0,194	81	72	1,12
smo	Lycopodiopsida	<i>Selaginella moellendorffii</i>	SELMODRAFT_178381	SELMODRAFT_450481	681	6,25	73156,6	30,26	95,32	-0,035	71	64	1,11
npu	Bacteria	<i>Nostoc punctiforme</i>	-	R2068	569	5,7	62244	29,74	98,86	0	54	43	1,26
aor	Endophyte	<i>Aspergillus oryzae</i>	AOR_1_122114	AOR_1_934174	713	6,7	77340,9	39,44	90,63	-0,12	66	62	1,06
ang	Endophyte	<i>Aspergillus niger</i>	ANI_1_106134	ANI_1_738114	744	5,72	81132,1	34,44	92,31	-0,069	83	65	1,28
afm	Endophyte	<i>Aspergillus fumigatus</i>	AFUA_6G04740	AFUA_2G09110	727	5,73	78920,6	39,53	93,23	-0,043	69	52	1,33
nfi	Endophyte	<i>Neosartorya fischeri</i>	NFIA_051290	NFIA_084640	594	5,39	64585,2	42,76	91,84	-0,052	59	41	1,44
act	Endophyte	<i>Aspergillus clavatus</i>	ACLA_095800	ACLA_080920	663	5,65	72418,5	35,53	96,06	-0,013	72	55	1,31
ure	Free Life	<i>Uncinocarpus reesii</i>	UREG_03045	UREG_04219	710	5,82	76773,9	37,23	97,04	0,024	66	53	1,24
val	Pathogen	<i>Verticillium albo-atrum</i>	VDBG_00158	VDBG_08166	675	6,52	72344,7	40,03	95,35	-0,037	62	57	1,09
cci	Free Life	<i>Coprinopsis cinerea</i>	CC1G_08232	CC1G_06838	733	6,24	78887,3	42,9	91,87	-0,143	73	61	1,20
uma	Pathogen	<i>Ustilago maydis</i>	UM06217.1	UM00078.1	724	6,1	79227,3	39,47	90,58	-0,172	79	71	1,11
mgr	Pathogen	<i>Magnaporthe oryzae</i>	MGG_03982	MGG_10036	626	6,95	66523,4	32,75	95,61	-0,047	61	60	1,02
pgr	Pathogen	<i>Puccinia graminis</i>	PGTG_05488	PGTG_10006	742	6,57	80660,6	42,32	104,53	0,035	66	61	1,08
Média					704	6,06	76396,8	34,89	92,53	-0,113	75,96	66,09	1,16
Desvio Padrão					40	0,32	4469,72	3,72	3,27	0,074	8,60	9,18	0,09

Legenda: NRes. - Comprimento em resíduos; GRAVY - *Grand Average Hydropathy Index*; pI - ponto isoeletrico; ININ - Índice de Instabilidade; AI - Índice Alifático; (-)/(+) - Razão de aminoácidos carregados negativamente pelos aminoácidos carregados positivamente.

O teste do relógio molecular foi efetuado por comparação do valor de ML para a topologia dos dados com e sem as restrições de relógio molecular sob o modelo LG (+F+G+I) (LE e GASCUEL, 2008). As diferenças nas taxas evolutivas entre locais foram modelados usando uma distribuição discreta Gama (G) de "n" parâmetros e permitiu a existência de sítios invariantes (I).

Várias características da proteína foram determinadas para o agrupamento filogenético e comparações. O ponto isoelétrico (pI) e massa molecular (MW) de cada sequência da proteína foram inferidos com a ferramenta *Compute pI/Mw* (BJELLQVIST *et al.*, 1993; BJELLQVIST *et al.*, 1994; WILKINS *et al.*, 1999), a variabilidade presente nas sequências foi acessada através do servidor PVS (<http://imed.med.ucm.es/PVS/>), e as principais características das proteínas como o índice de instabilidade (ININ), índice de resíduos alifáticos (AI), coeficiente médio de hidrofobicidade (GRAVY), Arg+Lys (R+K), Asp+Glu (D+E), razão (-)/(+), composição atômica (C, H, N, O e S) e as frequências de uso de monômeros (MUBs) foram avaliados utilizando-se a ferramenta *ProtParam Tool* (WILKINS *et al.*, 1999). Testes de normalidade da distribuição e de correlação de Spearman para dados não paramétricos com aproximação gaussiana com atribuição de valores de P, a 5% de significância entre os dados recolhidos foram feitos no *software* GraphPad Prisma Versão 5.0 (<http://www.graphpad.com/>). Uma análise de agrupamentos foi realizada no programa BioEstat 5.3 (AYRES *et al.*, 2007) usando a distância euclidiana dos parâmetros principais da proteína (comprimento, pI, ININ, GRAVY e (-)/(+)) com ligação completa e variáveis padronizadas. As características elementares e MUBs foram avaliados através de mapas de calor obtidos no *Heat Map Hierarchical Clustering* no servidor HIV (<http://www.hiv.lanl.gov/>) usando o método completo com distância euclidiana e a estabilidade dos *clusters* formados avaliada pelo método de *bootstrap* com 1000 pseudorreplacas. Todos os experimentos seguiram procedimentos previamente estabelecidos por Luiz e colaboradores (2013).

A fim de se investigar a presença de códons selecionados positivamente, realizou-se a estimativa da pressão seletiva sob cada resíduo da proteína, calculada a partir da proporção de substituições não-sinônimas por sinônimas (ω), como anteriormente descrito (COSTA *et al.*, 2010). Para tanto, as sequências proteicas foram convertidas em ORFs, por meio de retrotradução e normalização obtidos pelo servidor Protogene (MORETTI *et al.*, 2006) e pelo pacote Bioedit (HALL, 1999), utilizando o

código genético padrão, para normalização as frequências de uso de códons, tornando as comparações mais precisas. As análises, então, foram realizadas no servidor Selecton v.2.1 (DORON-FAIGENBOIM *et al.*, 2005; STERN *et al.*, 2007). A significância estatística do logaritmo da taxa de verossimilhança foi testada por meio da comparação dos valores obtidos nos cálculos com uso de dois modelos aninhados: um modelo nulo que não assume qualquer seleção (M8a) (SWANSON *et al.*, 2003) e um modelo alternativo que o faz (M8) (YANG *et al.*, 2000).

Muitas abordagens foram utilizadas para determinar a extensão da recombinação no conjunto de dados da PAL, uma vez que estudos anteriores (EMILIANI *et al.*, 2009; BAGAL *et al.*, 2011; 2012) já haviam sugerido tais fenômenos. O método GARD (KOSAKOVSKY POND *et al.*, 2006) disponível no servidor Datamonkey (DELPORTE *et al.*, 2010) foi usado para procurar evidências de incongruências filogenéticas e identificar o número e a localização dos pontos de interrupção correspondentes a eventos de recombinação. A fim de confirmar os resultados do GARD, foi realizado uma busca por pontos de interrupção (*breaking points*) pelos seguintes métodos implementados no programa RDP3 (MARTIN *et al.*, 2010): RDP (MARTIN e RYBICKI, 2000), GENECONV (PADIDAM *et al.*, 1999), BootScan (MARTIN *et al.*, 2005), Maximum Chisquared Test (MaxChi) (SMITH, 1992), Chimera (POSADA e CRANDALL, 2001), Siscan (GIBBS *et al.*, 2000) e 3Seq (BONI *et al.*, 2007). As análises foram realizadas com as configurações padrão para os métodos de detecção, e os resultados foram submetidos a um teste de Bonferroni corrigido, utilizando como *cut-off* a 5% de probabilidade, aliado à exigência de que cada evento em potencial devesse ser detectado simultaneamente por cinco ou mais métodos.

RESULTADOS E DISCUSSÃO.

Características da Proteína

Usamos dados da Tabela 4 para realizar nossas comparações entre as características da PAL (Tabela 4) e após os testes de normalidade dos dados, pudemos inferir que apenas o ponto isoelétrico (pI) e o coeficiente médio de hidrofobicidade (GRAVY) seguem a distribuição gaussiana ($P_{\text{teste de Kolmogorov-Smirnov}} > 0,05$; $P_{\text{teste D'Agostino}}$

& Pearson normalidade Omnibus > 0,001). Assim, nós avaliamos o teste de correlação para um conjunto de dados não-paramétricos, usando o teste de correlação de Spearman (Tabela 5), usando sistemas de pesos matemáticos para comparação entre grupos.

Tabela 5. Correlação entre as características da PAL. Os coeficientes de correlação de Spearman (r) e seus valores de P são mostrados, respectivamente, abaixo e acima da diagonal da tabela. Os valores em negrito são considerados significativos (P < 0,05). Grupos referem-se a natureza da sequência de plantas (peso 0), bactérias (peso 1) e fungos (peso 2). Os valores de correlação e P foram obtidos a partir do teste de correlação de Spearman, no *software* GraphPad Prism 5.

	NRes.	pI	MW	ININ	AI	GRAVY	D+E	R+K	(-)/(+)	Grupo
NRes.		0,589	3,7e-13	0,702	0,075	0,040	0,019	0,086	0,923	0,644
pI	0,099		0,506	0,791	0,209	0,268	0,320	0,168	4,8e-11	0,843
MW	0,912	0,121		0,649	0,027	0,002	0,029	0,151	0,751	0,686
ININ	0,070	-0,048	0,084		0,143	0,061	0,051	0,023	0,438	8,5e-05
AI	-0,319	-0,228	-0,391	0,264		1,8e-08	0,001	3,5e-05	0,325	0,004
GRAVY	-0,365	-0,201	-0,516	0,335	0,811		3,8E-04	4,3e-05	0,464	2,2e-04
D+E	0,410	-0,181	0,385	-0,347	-0,536	-0,590		1,4e-07	0,645	2,5e-05
R+K	0,308	0,249	0,259	-0,400	-0,663	-0,658	0,780		0,024	2,1e-07
(-)/(+)	0,018	-0,876	0,058	0,142	0,180	0,134	0,0846	-0,398		0,331
Grupo	-0,085	-0,036	-0,074	0,638	0,491	0,608	-0,672	-0,773	0,177	

As características apresentadas possuem diferentes níveis de correlação, e alguns são significativos. O comprimento de proteína pode ser correlacionado com MW (r = 0,912, p = 3,7x10⁻¹³), GRAVY (r = -0,365, p = 4,0x10⁻²) e D+E (r = 0,410, p = 1,96x10⁻²). Isto quer dizer que com o aumento do número de resíduos da PAL a MW tende a aumentar, juntamente com o teor de resíduos hidrofóbicos e carregados negativamente. O tamanho da proteína é altamente variável, com um desvio padrão de aproximadamente 40 resíduos. Assim, fragmentos presentes na sequência primária da enzima poderiam potencialmente ser descartados ao longo da história evolutiva, ou podem representar um processo de fixação. O principal fato dessa descoberta é que se não há efeitos nocivos sobre a redução do número de resíduos, a enzima pode ser projetada para fins industriais ou farmacêuticos, a fim de ter sua produção facilitada e sua atividade otimizada. Um fato que apoia as reduções não prejudiciais é a manutenção de atividade da PAL de *N. punctiforme*, que é funcional mesmo com 135 resíduos a menos do que a média.

O ponto isoelétrico foi correlacionado com a razão entre as cargas (-)/(+) da proteína ($r = -0,876$, $p = 4,8 \times 10^{-11}$), onde o acúmulo de aminoácidos carregados negativamente em relação a quantidade de aminoácidos carregados positivamente reduz o pI, o que torna a enzima mais ácida. Nossos dados aliados ao banco de dados de enzima Braunschweig - BRENDA (<http://www.brenda-enzymes.info/>) sugerem que a PAL desempenha a catálise preferencialmente em ambientes alcalinos, uma vez que seu pI se situa por volta de $6,0(\pm 0,3)$ e tem pH ótimo de 8 a 8,5.

O índice de instabilidade (ININ) revela a instabilidade da proteína num tubo de testes e é medido por meio da ocorrência de dipeptídeos específicos, que são frequentemente encontrados em proteínas instáveis. Assim, GURUPRASAD *et al.* (1990) estabeleceram que uma proteína cujo ININ é inferior a 40 pode ser classificada como estável. A proteína PAL apresentou um ININ médio de $34,89 (\pm 3,72)$, que a caracteriza como uma proteína estável. Essa característica foi negativamente correlacionada com a quantidade de aminoácidos carregados positivamente em PALs, R+K ($r = -0,400$, $p = 2,3 \times 10^{-2}$). Isto revela que existe uma pré-disposição de maiores estabilidades reforçadas pelo aumento do conteúdo de resíduos positivos. A razão de cargas pode ser correlacionada com valores de R+K ($r = -0,398$, $p = 2,4 \times 10^{-2}$) e essa variável também está relacionada a valores de D+E ($r = 0,785$, $p = 1,4 \times 10^{-7}$). Estas relações revelam uma inclinação à baixa variação de carga líquida na PAL, uma vez que se os valores de R+K são reduzidos, os valores de D+E são elevados. Uma vez que o momento de dipolo, gerado pelas α -hélices da PAL, é crucial para a eficiência catalítica a manutenção da carga líquida da PAL parece ser importante para estabilidade e função da enzima.

Os grupos de sequências de PAL de plantas (monocotiledôneas, dicotiledôneas e Lycopodiopsidae), bactérias e fungos (fungos endofíticos, de vida livre e agentes patogênicos) podem ser correlacionados com vários parâmetros da PAL, entretanto, vamos nos restringir as relações mais significantes. As monocotiledôneas (Monocots) são positivamente correlacionadas com os valores de R+K ($r = 0,435$, $p = 1,3 \times 10^{-2}$) e D+E ($r = 0,367$, $p = 3,9 \times 10^{-2}$), significando que há um aumento na quantidade de resíduos carregados, porém a proporção de cargas é mantida e não difere estatisticamente dos outros grupos. Com efeito, esta característica é muito importante para a enzima final, uma vez que sua atividade é dependente de interação dos aminoácidos carregados com o substrato e o cofactor MIO.

Os valores de MW ($r = 0,444$, $p = 1,1 \times 10^{-2}$), R+K ($r = 0,474$, $p = 6,0 \times 10^{-3}$) e D+E ($r = 0,448$, $p = 1,0 \times 10^{-2}$) estão correlacionados com o grupo de dicotiledôneas (Dicots). Esta relação implica em um maior comprimento que eleva o peso molecular juntamente com as quantidades de aminoácidos carregados. A razão de cargas é mantida entre os grupos, o que significa que a carga líquida da proteína se mantém. Os parâmetros ININ ($r = -0,396$, $p = 2,5 \times 10^{-2}$), AI ($r = -0,627$, $p = 1,2 \times 10^{-4}$) e GRAVY ($r = -0,817$, $p = 1,2 \times 10^{-8}$) também foram correlacionados com o grupo das dicotiledôneas.

Os valores de hidrofobicidade da PAL, medidos como GRAVY, apresentaram grandes variações, onde monocotiledôneas tiveram valores 54,03% inferiores aos das dicotiledôneas; enquanto o grupo das plantas teve um GRAVY equivalente a 161,26% da média apresentada pelos fungos. Isto significa que a PAL tem uma ordem de hidrofília decrescente sendo a PAL dos fungos, seguida pelas monocotiledôneas e depois dicotiledôneas. Tal resultado era esperado uma vez que as enzimas de fungos muitas vezes precisam ser secretadas para o meio ambiente e aumentos da hidrofílicidade são necessários, a fim de aumentar-se a eficiência da difusão. Dessa forma, o grupo das dicotiledôneas parece sofrer um aumento na estabilidade através de um aumento diferencial de cargas da cadeia do polipeptídeo, que também aumenta sua hidrofília (evidenciado pelos valores mais baixos de GRAVY e AI). No decurso de uma maior disponibilidade para o solvente, as enzimas deste grupo teriam melhor desempenho catalítico e menor índice de renovação. Devido a todas estas características, este grupo de enzimas provavelmente seria mais facilmente produzido sob condições recombinantes, como observado também por Gámez e colegas (2004).

Os endófitos apresentaram correlação com pI ($r = -0,368$, $p = 3,8 \times 10^{-2}$), R+K ($r = -0,482$, $p = 5,0 \times 10^{-3}$) e a razão de cargas ($r = 0,387$, $p = 2,9 \times 10^{-2}$). As PAL deste grupo tendem a ser mais ácidas. Na condição de endófitos, estes microrganismos entram em contato com um grande volume de compostos metabólicos celulares dos hospedeiros, vivendo em seus espaços intercelulares, na maioria dos casos, e em especial nos fungos, que tem nichos relacionados ao filoplano (ARAÚJO *et al.*, 2010), o que pode estar relacionado à variação do pH ambiental destas espécies. A redução da quantidade de resíduos carregados positivamente ao longo da sequência de proteína, induz a um aumento na razão de cargas, que instabiliza a proteína. A correlação estabelecida com ININ ($r = 0,433$, $p = 1,3 \times 10^{-2}$) mostra que estas enzimas são geralmente mais instáveis

do que a dos outros grupos, criando uma necessidade de maiores níveis de produção e ciclagem.

Os patógenos se correlacionam com o pI ($r = 0,451$, $p = 1,0 \times 10^{-2}$), D+E ($r = -0,412$, $p = 1,9 \times 10^{-2}$) e a razão de cargas ($r = -0,451$, $p = 1,0 \times 10^{-2}$). Como o grau de correlação revela, o conteúdo de aminoácidos de carga negativa tende a diminuir nestas enzimas, desta forma o pI tende a aumentar, enquanto a razão de cargas também diminui. Assim, o alto grau de hidrofilia poderia alterar a estabilização do *folding* protéico, uma vez que altos níveis de estabilidade foram correlacionados com um maior conteúdo de resíduos carregados distribuídos ao longo da PAL em proporções fixas. As implicações evolutivas dessas relações mostram que este grupo acumula enzimas alcalinas, e por conseguinte, como há uma mudança na carga total da enzima haverá, provavelmente, uma mudança em sua estabilidade. Este processo de alcalinização da PAL pode ser relacionado com o ambiente em que vivem os patógenos, adaptando-se aos resíduos metabólicos de seus hospedeiros, os quais tendem a ser ácidos devido à atividade celular oxidativa. Isto implica em uma evolução dirigida e estreita para adaptação de suas enzimas à evolução do hospedeiro, tornando sua vida dissociada, muito mais complexa no sentido da perda de valor adaptativo da enzima.

Na verdade, nos grupos testados as diferentes tendências para acumular e distribuir a carga líquida ao longo da proteína parecem ser cruciais. Isto, provavelmente, leva a um *folding* proteico diferencial, com parâmetros de reação distintos.

Os custos bioenergéticos de síntese de polímeros biológicos (DNA, RNA e proteínas) são diferentes em cada táxon devido ao uso diferencial dos diversos aminoácidos, devido ao seu custo associado, isto é, o número de átomos de C, N e S, necessários para construí-los (AKASHI e GOJOBORI, 2002; BRAGG e WAGNER, 2009; ELSER *et al.*, 2011). Dessa forma, a seleção natural gera um viés de uso de monômeros diferentes (MUBs) nesses polímeros, baseado não só nos custos bioenergéticos, mas também em sua composição elementar de acordo com a disponibilidade dos mesmos no ambiente.

Neste sentido, quando o conteúdo de aminoácidos ricos em N (Arg, Gln, Asn, Lys e His) foi avaliado no conjunto de dados da PAL (Figura 9), observou-se que as plantas têm 5,96% mais resíduos deste tipo do que os fungos. Já as monocotiledôneas têm em média 6,42% menos resíduos ricos em N, quando comparadas com

dicotiledôneas. Os resíduos ricos em C (Phe, His e Trp) foram distribuídos diferentemente ao longo dos táxons, e em média, as plantas têm 5,23% menos que os fungos, enquanto monocotiledôneas apresentaram um conteúdo 17,73% inferior ao das dicotiledôneas. As plantas ainda têm em média 1,93% menos resíduos ricos em O (Gln, Asn, Thr e Tyr) do que os fungos; e monocotiledôneas têm um conteúdo de aminoácidos ricos em O por volta de 9,56% inferior ao das dicotiledôneas.

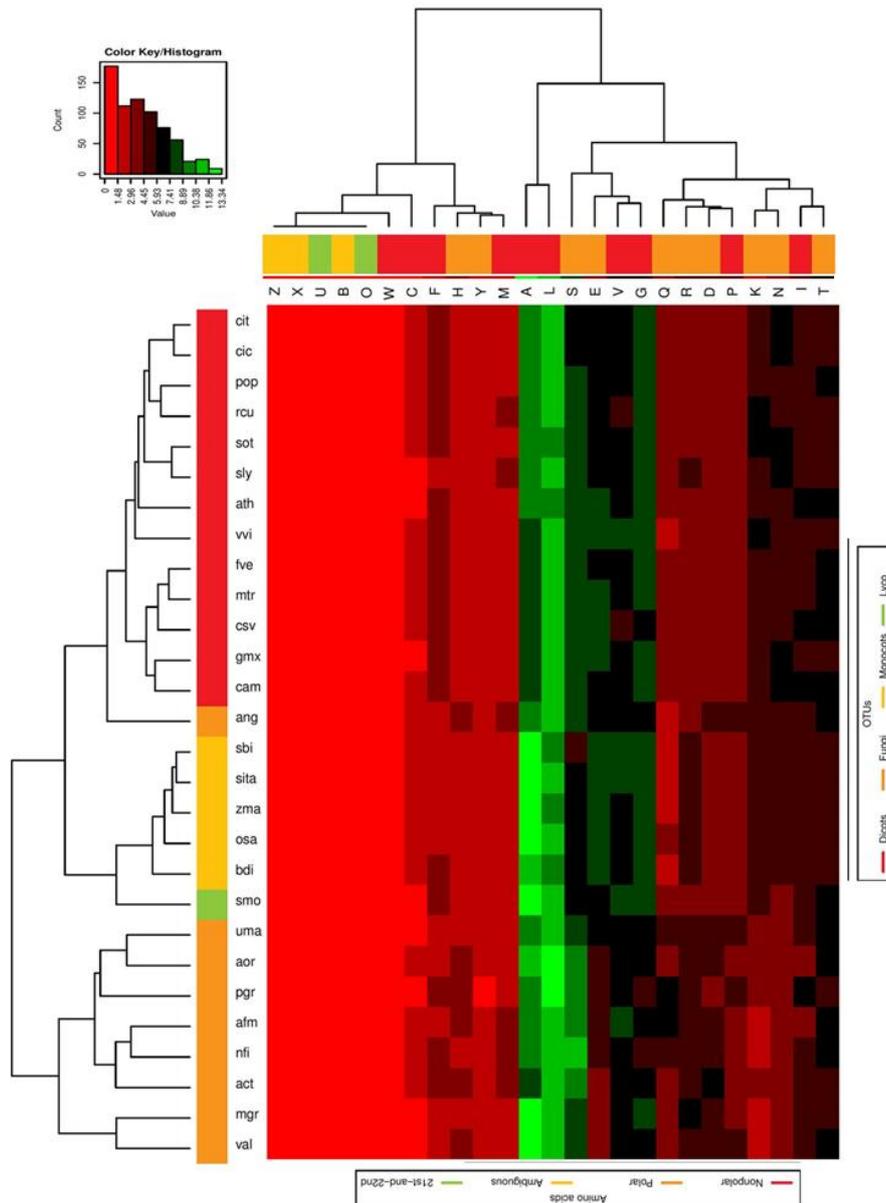


Figura 9. Desvios do uso de aminoácidos detectados no *set* de seqüências da proteína PAL. O agrupamento das amostras foi baseado no uso percentual de aminoácidos, tendo em vista o número total de resíduos. O mapa de calor mostra a

magnitude proporcional da mudança comparativa em cada grupo de amostras, de acordo com o gráfico-chave mostrado no canto esquerdo superior. O agrupamento das sequências é mostrado lateralmente e foi realizado pelo método de UPGMA. Os aminoácidos foram agrupados de acordo com a frequência de uso e são mostrados acima do mapa de calor.

Desta forma, parece que o N é o agente mais rigoroso na seleção da PAL, pois a correlação das Plantas ($r = -0,412$, $P_{\text{Spearman}} = 0,029$) e fungos ($r = 0,412$, $P_{\text{Spearman}} = 0,029$) com o teor de N mostra que seu uso sofreu um viés negativo. Esta relação inversa entre plantas e fungos pode explicar o papel da enzima no metabolismo desses organismos, como o catabolismo ou anabolismo de compostos nitrogenados. Em fungos, a PAL é predominantemente utilizada na reciclagem de carbono e nitrogênio, uma reação catabólica (FRITZ *et al.*, 1976); enquanto em plantas este mecanismo gera o ácido trans-cinâmico, usado na síntese de metabólitos secundários, um processo anabólico (CAMM e TOWERS, 1973; FRITZ *et al.*, 1976; TAIZ e ZEIGER, 2006). Realmente, há uma seleção ecofisiológica para conservação de nutrientes visando a maquinaria celular catabólica, expressa em resposta à baixa disponibilidade de nutrientes (ACQUISTI e KUMAR *et al.*, 2009). Por sua vez, a maquinaria anabólica implica em um alto custo nutricional durante o crescimento rápido promovido pela disponibilidade de nutrientes. Desta forma, o papel metabólico da PAL é essencial para determinação de sua arquitetura molecular, assim como de outras estruturas celulares de acordo com otimização de recursos e alocação de nutrientes (CARLSON, 2007; ACQUISTI e KUMAR *et al.*, 2009; ELSER *et al.*, 2011).

Ainda discutindo os teores de N na PAL, temos que o grupo das dicotiledôneas é mais caro em termos elementares do que as monocotiledôneas, provavelmente, porque sua natureza ancestral otimizou essas estruturas e provém de longos períodos de privação nutricional. Esta pressão diferencial ainda pode estar relacionada com o processo de domesticação, neste grupo outros autores encontraram sinais significativos de diminuição da conservação de N nos proteomas (ACQUISTI e ELSER *et al.*, 2009), representando um caso de evolução convergente na composição da PAL, causado pelo relaxamento das limitações de recursos em linhagens independentes, tal como proposto por Elser *et al.* (2011) para proteínas de plantas domesticadas de um modo geral.

Os fungos usam PAL em vias de degradação de compostos para obtenção de N e C, portanto, estas maquinarias devem utilizar menos N, pois são sintetizadas em momentos de depleção elementar. Os MUBs ainda foram eficientes em segregar unidades taxonômicas em grandes grupos, tendo uma resolução maior que a composição elementar (Tabela 6, Figura 9).

Tabela 6. Agrupamento diferencial da PAL por meio de diferentes abordagens.

Espécie	Bioquímica	MUBs	ML	Filogenia
<i>O. sativa</i>	I-A	I	I-A	BEP clade
<i>S. bicolor</i>	I-A	I	I-A	PACMAD clade
<i>S. italica</i>	I-A	I	I-A	PACMAD clade
<i>Z. mays</i>	I-A	I	I-A	PACMAD clade
<i>B. distachyon</i>	I-A	I	I-A	BEP clade
<i>S. moellendorffii</i>	I-A	I	-	Lycopodiopsidae
<i>R. communis</i>	I-B	II-A	I-B	Malphigales
<i>P. trichocarpa</i>	I-B	II-A	I-B	Malphigales
<i>S. lycopersicum</i>	I-B	II-A	I-B	Asterids
<i>S. tuberosum</i>	I-B	II-A	I-B	Asterids
<i>A. thaliana</i>	I-B	II-A	I-B	Brassicales
<i>G. max</i>	I-B	II-B	I-B	Fabales
<i>C. arietinum</i>	I-B	II-B	I-B	Fabales
<i>F. vesca</i>	I-B	II-B	I-B	Rosales
<i>C. sinensis</i>	I-B	II-A	I-B	Sapindales
<i>C. clementina</i>	I-B	II-A	I-B	Sapindales
<i>V. vinifera</i>	I-B	II-B	I-B	Vitaceae
<i>M. truncatula</i>	I-C	II-B	I-B	Fabales
<i>C. sativus</i>	I-C	II-B	I-B	Cucurbitales
<i>U. maydis</i>	I-C	III-A	III	Basidiomycota
<i>A. oryzae</i>	II-A	III-A	II	Eurofiales
<i>M. oryzae</i>	II-A	III-B	III	Sordariomycetes
<i>V. albo-atrum</i>	II-B	III-B	II	Sordariomycetes
<i>P. graminis</i>	II-B	III-A	-	Basidiomycota
<i>A. niger</i>	III	-	II	Eurofiales
<i>A. fumigatus</i>	III	III-A	II	Eurofiales
<i>N. fischeri</i>	III	III-A	II	Eurofiales
<i>A. clavatus</i>	III	III-A	II	Eurofiales

A análise de agrupamentos (Figura 10, Tabela 6) reuniu a um nível de 50% de similaridade três grupos bioquimicamente relacionados de PAL (I, II e III). O grupo I é

predominantemente composto por plantas, onde podemos observar a Lycopodiopsidae (*S. moellendorffii*) muito mais semelhante das monocotiledôneas do que das outras sequências e um agrupamento da sequência de *O. sativa*, um membro do clado BEP, com os membros do clado PACMAD (*S. italica*, *Z. mays*, *S. bicolor*) sendo mais relacionado com a sequência de *S. italica*. Isto mostra que abordagens multifásicas mostram resultados diferentes dos convencionais. Observa-se que as Malphigiales (*R. communis* e *P. trichocarpa*) foram separadas e agrupadas com uma asterídea (*S. tuberosum*) e um membro do grupo Rosales (*F. vesca*). As asterídeas (*S. tuberosum* e *S. lycopersicum*) foram segregadas, apesar de filogeneticamente comporem um grupo isolado. A semelhança entre *C. arietinum* e *V. vinifera* é igualmente surpreendente, uma vez que seus grupos Magnoliopsidae (Fabales) e Vitaceae, respectivamente, são filogeneticamente distantes, mas bioquimicamente parecem próximos.

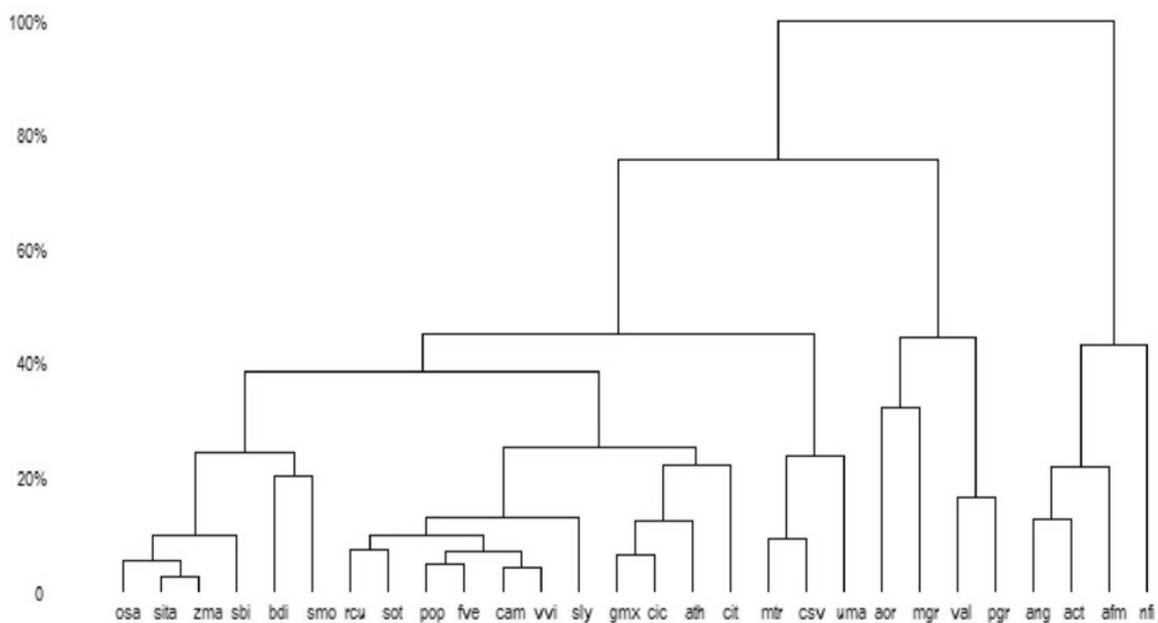


Figura 10. Análise de agrupamento de táxons de acordo com as principais características bioquímicas da PAL. A análise foi feita utilizando-se a distância Euclidiana, com ligação completa como método de agregação e variáveis padronizadas (número de resíduos, pI, ININ, GRAVY e razão de cargas) no *software* BioEstat v.5. O dendrograma é baseado na similitude das sequências e as distâncias foram normalizadas utilizando-se a maior distância calculada.

O membro do grupo Fabales (*G. max*) foi agrupado com o representante dos Sapindales, *C. clementina*, que está próximo do outro membro de seu grupo, *C. sinensis*. Dessa forma *A. thaliana* (Brassicales) que é membro do grupo das Malvideas, também foi agrupado entre eles. Neste grupo ainda houve uma incongruência final, o agrupamento de *M. truncatula* (Fabales) e *C. sativus* (Cucurbitales), ambos Magnoliopsidae, que foram agrupados juntamente com um Basidiomycota (*U. maydis*). Apesar disso, a análise mostrou uma tendência de agrupamento de sequências específica, onde culturas clássicas de cultivo foram reunidas, provavelmente devido ao seu processo de evolução que sofreu influência antrópica e uma potencial acumulação de características semelhantes.

A razão N/C de uma proteína expressa como os conteúdos de N e C estão mudando em função de si mesmos. A correlação apresentada entre a razão N/C da PAL e dos grupos de monocotiledôneas ($r = 0,375$, $P_{\text{Spearman}} = 0,0049$) e dicotiledôneas ($r = -0,510$, $P_{\text{Spearman}} = 0,006$) apenas reforça a idéia do papel da domesticação de culturas na evolução, como sugerido também por estudos anteriores (BRAGG e HYDER 2004; ACQUISTI e ELSER *et al.*, 2009).

O grupo II é mais homogêneo e é composto predominantemente por membros Ascomycotas, com uma alta representação de Sordariomycetes (*M. oryzae* e *V. albo-atrum*) agrupados diferencialmente com um membro dos Eurofiales (*A. oryzae*) e um Basidiomycota (*P. graminis*), respectivamente. Finalmente, o grupo III não possui inconsistências taxonômicas, onde apenas sequências de Eurofiales foram agrupadas (*A. niger*, *A. clavatus*, *N. fisheri*, *A. fumigatus*). Isto mostra que estes grupos, além de bioquimicamente relacionados não tem variação filogenética intrínseca, podendo refletir aspectos adaptativos do grupo devido à ancestralidade comum. A semelhança bioquímica mostra ainda como a pressão seletiva está atuando e como estas características poderiam representar uma força dentro da seleção natural.

Inferências filogenéticas

A matriz de substituição mais viável para a substituição de aminoácidos em ambos os casos, tanto para PAL e actina, foi o modelo LG (+G+I+F) (LE e GASCUEL, 2008), determinado para a PAL com os melhores critérios de informação Bayesiano

(20146,89) e de Akaike com o valor corrigido (19677,88) e ln L de -9775,62. A não uniformidade das taxas evolutivas entre os sítios pode ser modelada utilizando-se uma distribuição gama discreta (+G) com 5 categorias de variação, assumindo-se que uma certa fração dos sítios seja evolutivamente invariável (+I).

No alinhamento realizado com os G-blocos das proteínas PAL curadas identificou-se as regiões de resíduos conservados contíguos. A estimativa de sítios invariáveis encontrados foi de 74,13% das posições, indicando uma enzima relativamente conservada. A análise de variação (Figura 11) mostrou que valores mais altos para índice de variabilidade de Wu-Kabat foram encontrados em regiões contíguas, indicando a existência de blocos conservados. Esses blocos provavelmente tem relação com a manutenção do *fold*ing, interações com ligantes e ação catalítica. A variabilidade é concentrada em regiões com propensão à formação de *gamma-turns* e *beta-sheets*, e no início e fim de α -hélices.

As matrizes de distância para proteínas PAL e actina fornecidas pela análise de divergência evolutiva em rede (*Net Evolutionary Divergence*) entre grupos de sequências, pelo modelo de correção de *Poisson* (ZUCKERKANDL e PAULING, 1965) é mostrado na Tabela 7. Quando comparados entre eles, os grupos de PAL são em média mais distantes, em consonância com o esperado, uma vez que a actina é uma proteína relativamente conservada ao longo de vários táxons divergentes, devido seu papel estrutural (HELGASON *et al.*, 2003; SIN *et al.*, 2007; ZHANG *et al.*, 2010; LAHR *et al.*, 2011). Assim, a distância ainda revelou que os grupos de plantas são mais próximos entre si, que entre os outros. O grupo Lycopodiopsida é duas vezes mais distante das monocotiledôneas ($d_{actina} = 0,041 \pm 0,011$; $d_{PAL} = 0,236 \pm 0,025$) e do grupo das dicotiledôneas ($d_{actina} = 0,024 \pm 0,007$; $d_{PAL} = 0,207 \pm 0,022$), do que em comparação da distância entre eles ($d_{actina} = 0,014 \pm 0,005$; $d_{PAL} = 0,102 \pm 0,013$), pois eles sofrem uma separação daquele táxon enquanto a grande classe taxonômica Spermatopsida.

Os patógenos mostraram uma relação estrita com o grupo dos endófitos ($d_{actina} = 0,015 \pm 0,003$; $d_{PAL} = 0,144 \pm 0,019$), igualmente esperado, uma vez que esta nomenclatura é apenas uma classificação arbitrária, e quando à filogenia tal grupo é monofilético. Essa discussão traz a amplitude aparente das relações estabelecidas entre hospedeiro e parasita/comensal, quando em ambas as situações, os organismos devem desenvolver mecanismos moleculares para seu estabelecimento. A PAL poderia ter um papel duplo de estabelecimento e proteção, o que provavelmente torna tão interessante

sua manutenção e evolução direcionada, enquanto uma simples proteína desses organismos. O grupo de organismos de vida livre mostrou também uma pequena distância com os patógenos ($d_{actina} = -0,019 \pm 0,007$; $d_{PAL} = -0,038 \pm 0,021$).

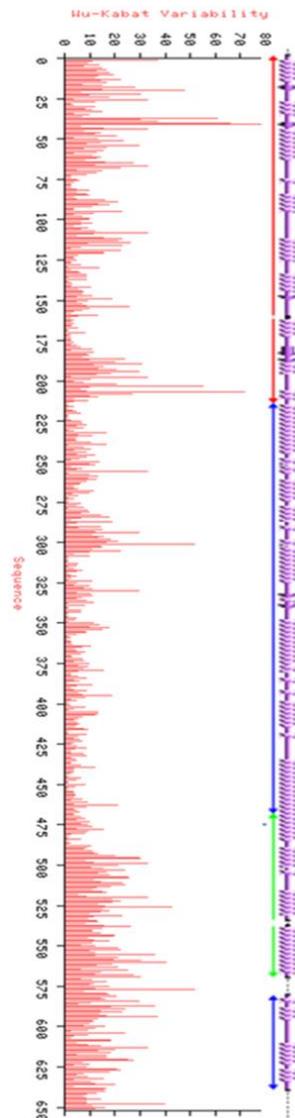


Figura 11. Variabilidade da proteína PAL. Os valores de variabilidade medidos com o índice de Wu-Kabat resíduo a resíduo são mostrados com a estrutura de proteína secundária. Setas indicam seções de folhas betas, espirais indicam α -hélices e segmentos lineares indicam estruturas desordenadas como *turns*. Os motifs da proteína são mostrados em diferentes cores.

As distâncias são maiores entre as bactérias e os outros grupos, devido à sua diferença característica da estrutura enzimática, em que procariotas não possuem um domínio de aproximadamente 120 resíduos de comprimento C-terminal multi-hélice, (RITTER e SCHULZ, 2004; WANG *et al.*, 2008; MOFFITT *et al.*, 2007; CALABRESE *et al.*, 2004; WANG *et al.*, 2005). Essa divergência na magnitude das distâncias entre PAL e actina sugere que em todos os casos há uma diferença na taxa de evolução delas. E ainda indica que os diferentes grupos provavelmente têm diferentes tipos de pressões seletivas. Em bactérias, geralmente esta enzima funciona ainda na fixação de nitrogênio, permitindo-lhes estabelecer várias relações ecológicas (ESTABROOK e SENGUPTA-GOPALAN, 1991) com outros organismos, mais especificamente em *Nostoc* sp. esta enzima é agrupada com genes do metabolismo secundário, não como observado para os outros (MOFFITT *et al.*, 2007), reduzindo a pressão sob esta proteína.

Tabela 7. Estimativas de divergência evolutiva entre grupos de sequências. A estimativa líquida média do número de substituições de aminoácidos por sítio entre os grupos de sequências é mostrada. A estimativa de erro padrão está mostrado acima da diagonal da tabela entre [], os dados faltantes são mostrados como "-".

ACTINA							
	Monocot	Dicot	Lycopodiopsidae	Bactéria	Endófitos	Vida Livre	Patógenos
Monocot		[0,005]	[0,011]	[-]	[0,024]	[0,018]	[0,017]
Dicot	0,014		[0,007]	[-]	[0,024]	[0,017]	[0,017]
Lycopodiopsidae	0,041	0,024		[-]	[0,024]	[0,018]	[0,017]
Bactéria	-	-	-		[-]	[-]	[-]
Endófitos	0,161	0,154	0,159	-		[0,006]	[0,003]
Vida Livre	0,083	0,079	0,082	-	0,010		[0,007]
Patógenos	0,104	0,097	0,101	-	0,015	-0,019	

PAL							
	Monocot	Dicot	Lycopodiopsidae	Bactéria	Endófitos	Vida Livre	Patógenos
Monocot		[0,013]	[0,025]	[0,075]	[0,055]	[0,057]	[0,044]
Dicot	0,102		[0,022]	[0,076]	[0,053]	[0,053]	[0,043]
Lycopodiopsidae	0,236	0,207		[0,078]	[0,058]	[0,055]	[0,044]
Bactéria	1,036	1,039	1,089		[0,069]	[0,069]	[0,062]
Endófitos	0,698	0,652	0,717	0,949		[0,018]	[0,019]
Vida Livre	0,603	0,525	0,569	0,828	0,042		[0,021]
Patógenos	0,523	0,485	0,505	0,810	0,144	-0,038	

Conseguimos identificar algumas incongruências topológicas presentes na organização dos ramos da árvore ML da PAL (Figura 12), quando comparados a classificação filogenética disponível no KEGG e no *Tree of Life* da Web, mas nenhuma diferença pode ser visualizada na comparação das árvores ML de sequências PAL completas em comparação com a obtida dos G-blocos curados (dados não mostrados). Em todas as topologias, 3 grupos foram bem separados com um *bootstrap* de 100% (bactérias, fungos e plantas). Isso mostra que nestes grupos de organismos o gene da PAL está sob diferentes pressões seletivas, e ao longo dos eventos evolutivos, a PAL tem alcançado níveis diferentes de valores adaptativos.

Dentro do grupo de plantas, a separação da unidade taxonômica operacional (OTU) representante do grupo Lycopodiopsidae, para com os outros (*bootstrap* de 100) e sua conservação nas árvores de sequências curadas e não curadas tornou-se conclusiva, mostrando sua natureza ancestral. A distinção de monocotiledôneas e dicotiledôneas também é muito evidente, e as reais incongruências topológicas estão dentro desses grupos. O primeiro evento topológico perceptível identificado é a separação errônea entre os clados BEP e PACMAD da família Poaceae, em que a PAL tem um comportamento parafilético.

Em relação às dicotiledôneas, a separação de clados ocorre ainda a nível de família, a sequência da família Vitaceae é separada dos grandes grupos monofiléticos de rosídeas e asterídeas, com um nó suportado pela estatística (*bootstrap* de 99). As asterídeas, representadas pelo gênero *Solanum*, são distantes dos outros grupos e formam uma unidade taxonômica fechada, quando analisados contra outras sequências. Sua proximidade com outras dicotiledôneas revela uma boa organização de ramos quanto à classificação canônica (*bootstrap* de 100). As Magnoliopsidae foram separadas em dois grupos principais (Fabidae e Malvidae). As Malvidae (aqui representado por *Citrus* sp. e *A. thaliana*) foram segregadas em Sapindales (*Citrus* sp.) e Malvales (*A. thaliana*), sendo o último agrupado com as Malpighiales (*R. communis* e *P. trichocarpa*) com um forte suporte estatístico deste nó (*bootstrap* de 100 e 70, respectivamente).

As Fabidae foram separadas em dois grupos principais, Rosales (*F. vesca*) e uma segunda unidade taxonômica parafilética, composta por diferentes clados. Esta segregação não foi apoiada por estatísticas, evidenciando que estes grupos poderiam

ocupar um mesmo grau de separação. O segundo grupo, composto por Mapighiales, Fabales (*G. max*, *M. truncatula* e *C. arietinum*) e Cucurbitales (*C. sativus*) dividiram-se em nós individuais bem suportados. Apesar da particionamento errôneo das Malvidae, não foi detectada nenhuma incongruência topológica na classificação dos ramos a este nível.

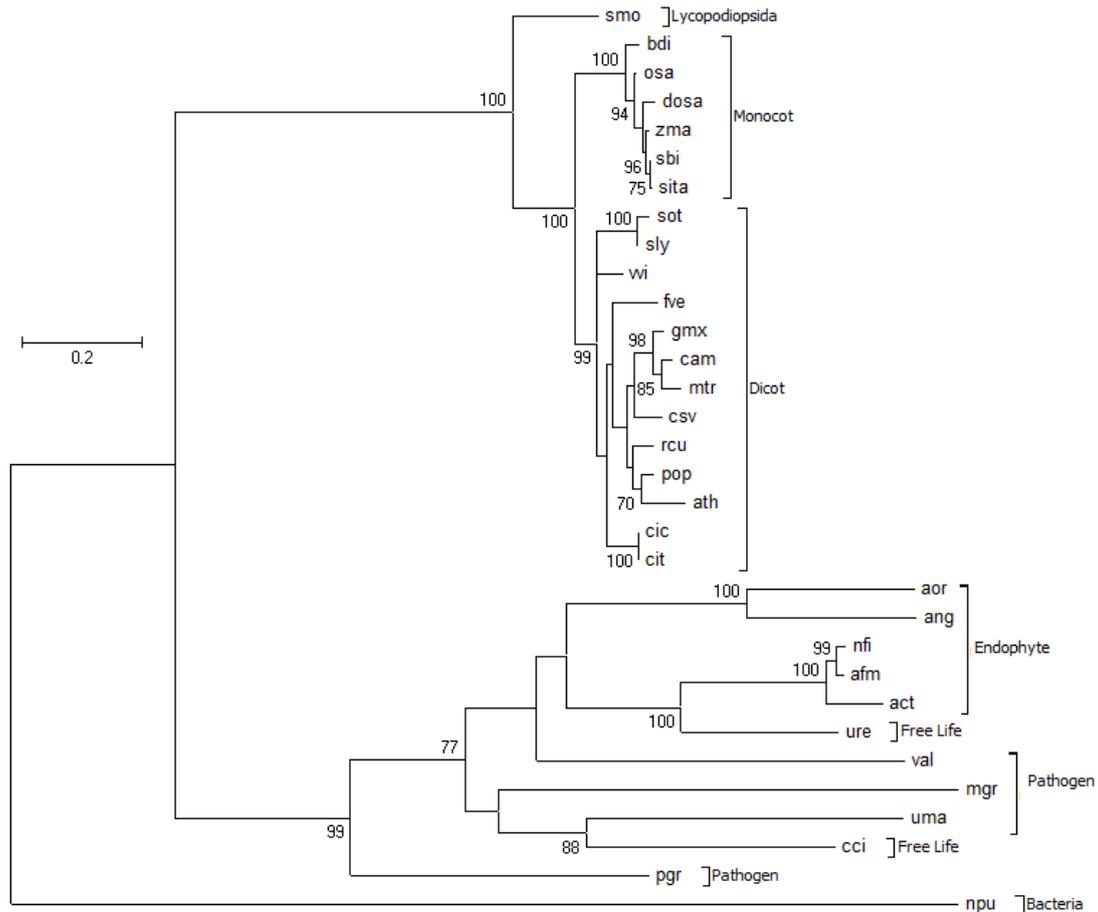


Figura 12. Reconstrução filogenética da história evolutiva da PAL pelo método ML. A árvore foi desenhada em escala, o comprimento dos ramos equivale ao número de substituições por sítio. Utilizou-se o modelo de Le e Gascuel (2008), com uma distribuição Gama com 5 parâmetros e *Bootstrapping* com 1000 pseudoreplicatas. A análise envolveu 32 sequências de aminoácidos. Todas as posições que contêm lacunas foram eliminadas da análise.

A filogenia dos fungos é conhecida e sua classificação em nossas árvores ML da PAL foram de compreensão complexa. Não houve um primeiro grau discriminatório

ao nível de filo, os Basidiomycota e Ascomycota, por exemplo. Em vez disso, o grupo dos basidiomicetos foi separado com um limite de confiança forte (*bootstrap* de 99) em um *singleton* (*P. graminis*) e um subagrupamento das PAL de *U. maydis* e *C. cinerea* juntamente com a de *M. oryzae* (Sordariomycete). Em nossa árvore ML, as unidades taxonômicas operacionais *A. oryzae* e *A. niger* excluem-se do núcleo do táxon Eurotiales, mas formam um grupo monofilético com outras sequências deste grupo, sob forte apoio estatístico (*bootstrap* de 100). A íntima relação estabelecida entre os endofíticos do gênero *Aspergillus* e *U. reesii* é explicada por sua filogenia, em que ocupam a mesma subclasse Eurotiomycetidae, sendo sua separação bastante clara quanto ao grau de ordem (Eurotiales e Onygenales, respectivamente).

Evolução molecular

Através de uma análise do gráfico de *Splits* (Figura 13) foi possível identificar a separação clara de Lycopodiopsidae, eudicotiledôneas e monocotiledôneas, reforçando o observado na árvore ML (Figura 12). Muitos eventos podem trazer à reticulação observada, indicando que esses eventos são recentes na história biológica, desde que Lycopodiopsidae é acreditado ser mais antigo do que os outros grupos de plantas e se segrega antes deles. Essa configuração ainda indica que, provavelmente, eventos como recombinação devem estar presentes.

Os fungos apresentam um menor grau de reticulação, que torna sua história evolutiva mais simples. Um evento de reticulação é observado, segregando as OTUs *A. oryzae* e *A. niger* pertencentes ao grupo dos Eurotiales, que são mais estreitamente relacionados com os Onygenales, nomeadamente *U. reesii*. Analisando-se a distribuição, a partir de sua classificação quanto ao papel ecológico (aqui adotada), podemos ver uma distinção na distribuição dos patógenos, onde *V. albo-atrum* (Sordariomycete) está mais próximo dos Eurotiomycetes do que de seu próprio grupo clássico. O grupo das OTUs de Basidiomycetes não apresenta nenhum *cluster* restrito e se dispersa sob uma malha de reticulação aberta.

O teste do relógio molecular (Tabela 8) provou que os grupos de sequências testados de PAL não evoluem no mesmo ritmo ($P = 0$), enquanto parece que o mesmo

não ocorre com a actina ($P > 0,05$). Estes dados ainda trazem algumas perguntas sobre como este processo esta ocorrendo e se somente pressões seletivas estão agindo, ou se outros processos, como a recombinação, poderiam estar atuando, no sentido de se gerar diversidade. A conclusão principal foi que a reticulação aliada aos resultados do relógio molecular revelou que um conjunto diferente de táxons está evoluindo a taxas diferenciais, com padrões heterogêneos para esta enzima. Sendo assim, várias hipóteses foram testadas para discutir o assunto.

Tabela 8. Resultados do teste de relógios moleculares utilizando o método de máxima verossimilhança (ML). A hipótese nula de que a taxa evolutiva é igual em toda a árvore foi rejeitada ao nível de significância de 5% para PAL ($P = 0$), e não rejeitada para Actina ($P > 0,05$).

Parâmetros do relógio molecular para a proteína Actina				
	ln L	Parâmetros	(+G)	(+I)
Com relógio	-58738,210	51	0,782	0,47
Sem relógio	-1812,559	80	0,70	0,40
Parâmetros do relógio molecular para a proteína PAL				
	ln L	Parâmetros	(+G)	(+I)
Com relógio	-32505,621	52	1,231	0,09
Sem relógio	-9748,775	82	1,27	0,09

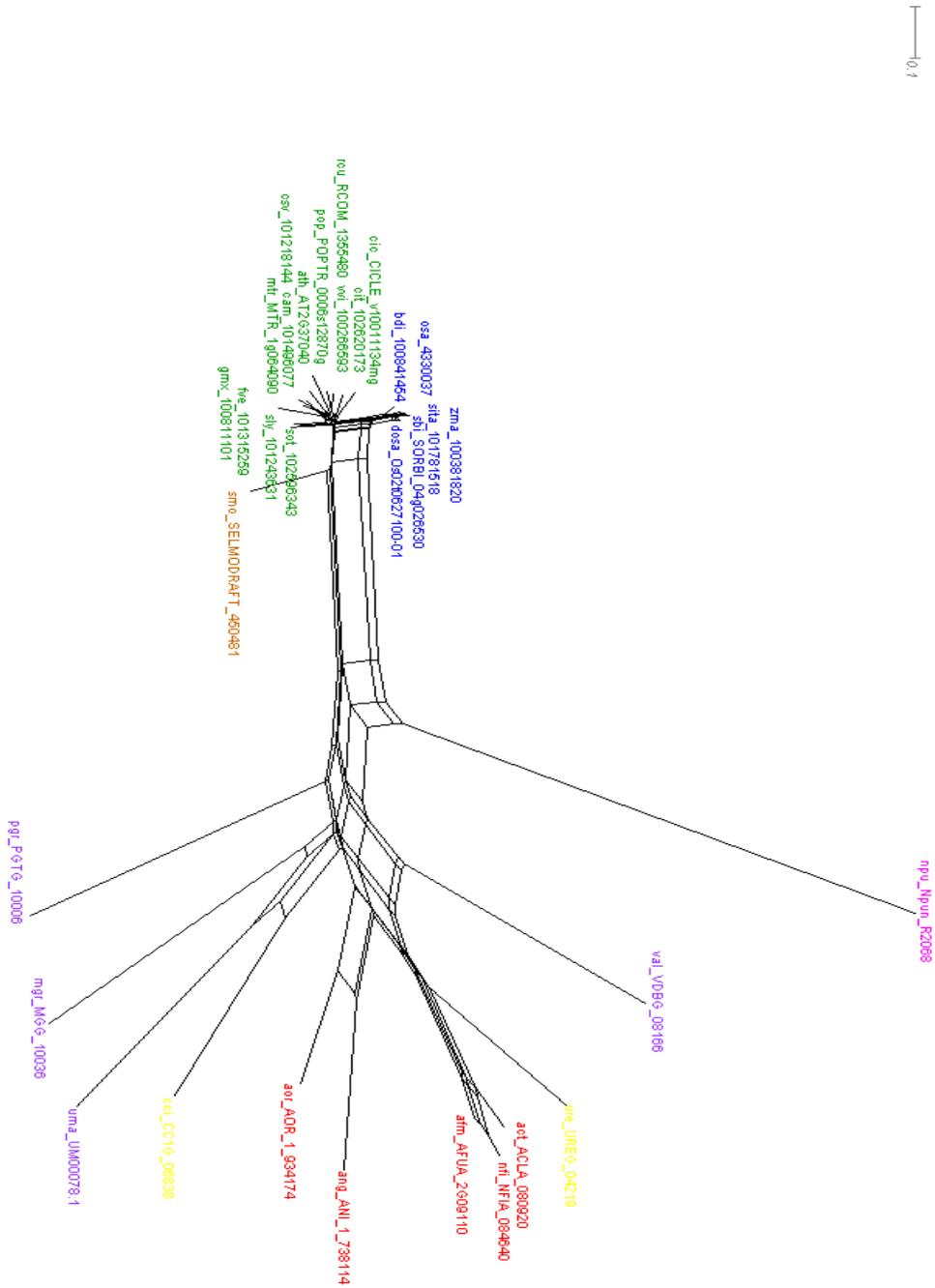


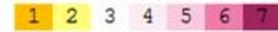
Figura 13. Gráfico de *Splits* da PAL completa. Os sítios de não informativos para parsimônia, lacunas e sítios conservados foram excluídos da análise. As OTUs em azul representam as monocotiledôneas, em verde representam as dicotiledôneas, em marrom representam as Lycopodiopsidae, as OTUs em vermelho representam fungos endofíticos, em roxo representam os patógenos, em amarelo representam os fungos de vida livre e em rosa representam as bactérias.

As forças de seleção positiva operando em cada códon de cada sequência de PAL foram evidenciadas pelo algoritmo Selecton e são mostrados como uma sequência única (Figura 14), que mostrou ser significativa ($p < 0,001$) com um ΔLnL de 515,1, este valor é referente à discrepância nos logaritmos de probabilidade obtida em dois modelos aninhados: um modelo nulo que não assume nenhuma seleção (M8a) (SWANSON *et al.*, 2003) e um modelo alternativo que o faz (M8) (YANG *et al.*, 2000). Aproximadamente 21,74% de resíduos da sequência consenso estão sob forte seleção positiva e eles tendem a se agrupar nos primeiros 100 resíduos e nos últimos 182 resíduos. As extremidades da proteína PAL estão sob uma evolução que a remodela, provavelmente para aumentar a eficiência do seu *folding* final, concentrando os resíduos catalíticos no interior da enzima. Essa ideia é reforçada pela concentração de resíduos sob forte seleção purificadora e seleção neutra nos resíduos intermediários da sequência (106° - 523°). Estes fragmentos com seleção diferencial poderiam configurar blocos que em eventos recombinatoriais ancestrais foram trocados entre organismos ortólogos e homólogos, significando que podem haver *cores* que sustentam a atividade catalítica.

A análise de recombinação é mostrada na Tabela 9 e Figura 15. A matriz de MaxChi (*Maximum Chi-squared test*) para cada potencial evento recombinante de 1 a 3 são mostrados na Figura 15 (respectivamente A a C). A relação entre os valores de P de qui-quadrado máximo de cada par possível de pontos de interrupção (BKP) e o comprimento correspondente dos domínios foi avaliada por um mapa de calor do total de pontos dos dados (Figuras 5. A a C). Os "picos" nos gráficos exibidos (regiões com concentração de cores amarelas a vermelho) são os *hot spots* para a recombinação. Isso foi validado através de um teste de Bonferroni a 5% de significância para cada ponto de borda entre as colunas (arbitrariamente pré-definido como o comprimento de um domínio baseado na categorização do mapa de calor) para verificar a significância da diferença entre as distribuições à esquerda e à direita desse ponto. Estes padrões de sítios polimórficos indica que a posição real do ponto de interrupção está localizada em algum lugar dentro de uma gama de sítios idênticos (CHAN *et al.*, 2009; MARTIN, 2009).

Legend:

The selection scale:



Positive selection

Purifying selection



Figura 14. Seleção positiva operando em cada códon da sequência de PAL. As sequências da Tabela 4 foram utilizadas para determinação das taxas de seleção operando em cada códon medidas no servidor Selecton por meio do fator ω , dado pela razão entre as taxas de mutação não sinônima por sinônima, e mostrados segundo a chave de cores da figura. A sequência mostrada é uma resultante da análise conjunta de todas as PAL do *dataset*.

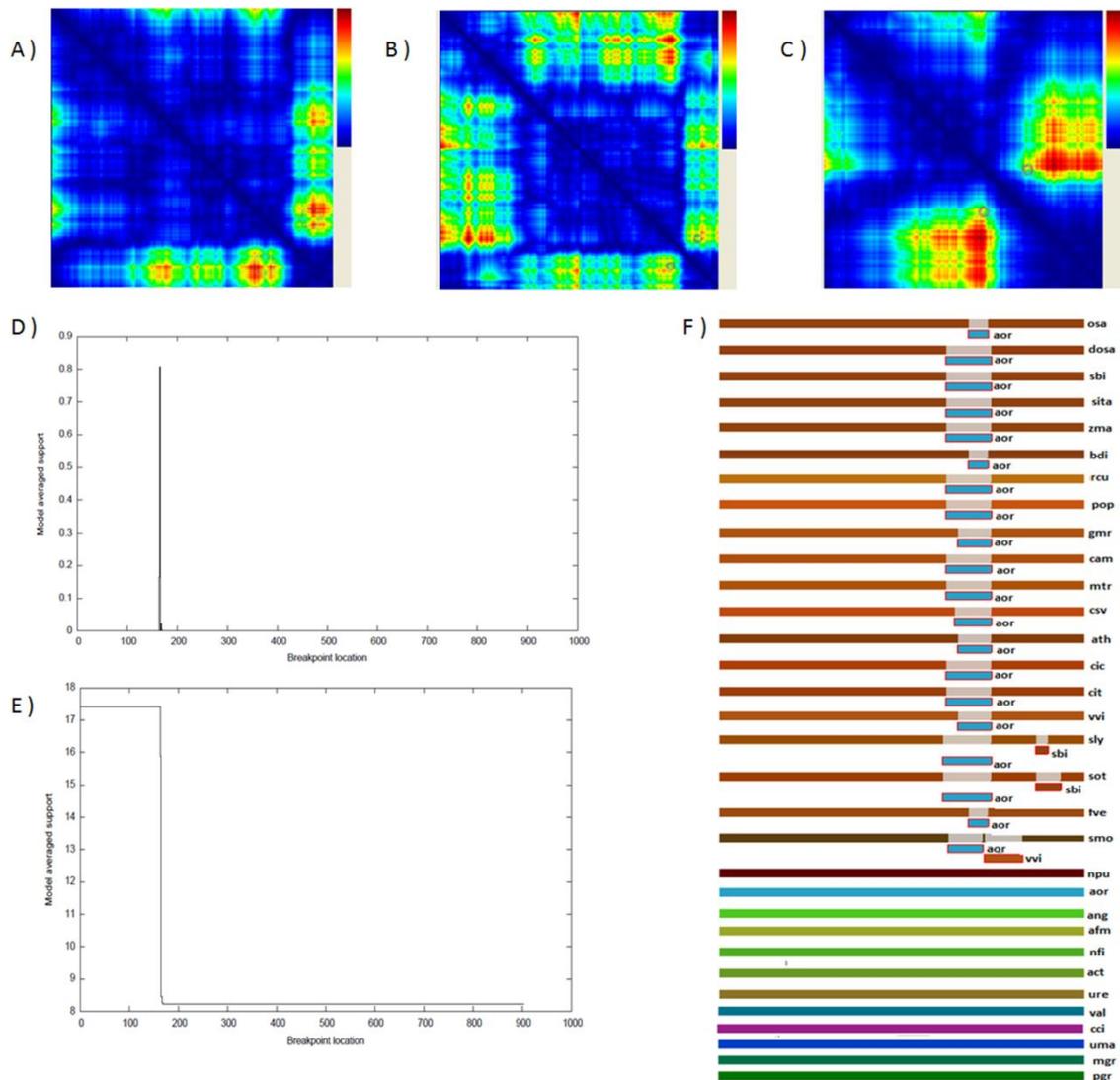


Figura 15. Eventos recombinantes envolvidos na evolução da PAL. As matrizes de MaxChi para cada potencial evento recombinante (1, 2 e 3) estão mostradas para os parentais e o recombinante principal (respectivamente de A a C), juntamente com os pontos de interrupção mais prováveis detectados pelo algoritmo do GARD (D) e a varredura do comprimento da árvore (E), ambos em resíduos de Proteína vs. Probabilidade. Cada sequência está representada por uma cor e o evento de recombinação está evidenciado pelo doador abaixo da sequência (F).

Comparando os valores do critério de Akaike corrigido (AICc) das melhores montagens do algoritmo GARD, obtivemos diferentes topologias entre segmentos c (47578,6) e o modelo que assume a mesma árvore para todas as partições inferidas pelo GARD na mesma árvore, mas permite comprimentos de ramos de diferentes partições (47621), sugerindo que, o modelo de árvore múltipla pode ser preferido em relação ao modelo de árvore isolada por uma taxa de provas de 100 ou mais, havendo pelo menos um BKP, o que por sua vez, reflete uma incongruência topológica significativa.

Na verdade, na análise do GARD, encontramos apenas um ponto de ruptura no resíduo 164, apoiado com um valor de P do teste KH à esquerda do BKP de 0,0002 e um valor de P para o teste à direita do BKP de 0,01. O algoritmo GARD ainda apresentou gráficos apoiando um provável ponto de ruptura encontrado em resíduos *versus* probabilidade (Figura 15. D e E), que reforça os dados estatísticos apresentados. O ponto de ruptura do GARD é consistente com os *hot spots* apresentados nas matrizes de MaxChi, mas os algoritmos implementados no RDP3 encontraram outro ponto mais provável, que foi validado por vários métodos diferentes. Os eventos recombinatórios são uma realidade na proteína PAL, já observado por Emiliani e colaboradores (2009) sob a ótica das transferências gênicas horizontais, mas a consistência dos resultados obtidos por diferentes algoritmos sugere ainda que esses eventos podem não ter sido completamente identificados e/ou compreendidos, devido à sua natureza complexa.

Cada evento recombinante (Tabela 9) e seu resumo gráfico (Figura 15. F) foram analisados individualmente. O evento recombinante 1 (valor de P_{BootScan} de $2,5 \times 10^{-5}$) parece ser onipresente em plantas (monocotiledôneas, dicotiledôneas e Lycopodiopsidae) e pode se referir a uma antiga realidade evolutiva de troca genética entre fungos e plantas, onde um parental ancestral doou o fragmento que foi conservado em *Aspergillus oryzae* (aor), seu descendente. A distribuição generalizada desse fungo, provavelmente facilitou a dinâmica dos eventos de recombinação (ROKAS, 2009). De fato, parece que este evento foi essencial para a estrutura da PAL nas plantas, uma vez que foi mantida em todas. Eventos que envolvem transferência horizontal de genes entre plantas e microrganismos são clássicos e parecem ser muito comuns ao longo da história evolutiva, assim como ocorre entre bactérias endofíticas e seus hospedeiros. Um caso interessante é dos fragmentos genômicos de *Ricinus communis* que foram conservados com níveis de 90% de identidade com o genoma de *Herbaspirillum*

seropedicae (PEDROSA *et al.*, 2011). Além disso, esse evento de transferência reforça a hipótese previamente apresentada por Emiliani *et al.* (2009), mostrada na Figura 7.

As asterídeas, gênero *Solanum*, podem ter sido segregadas na árvore ML da PAL (Figura 12) devido ao evento recombinante 2 (valor de P_{BootScan} de $3,1 \times 10^{-6}$). Desta forma, *Sorghum bicolor* foi doador, provavelmente por intermédio de um vírus em comum, ou alternativamente, por um endófito bacteriano ou fúngico comum. Propomos um cenário dependente de um vetor viral, devido à sua maior propensão de carrear material exógeno. Ainda supomos que provavelmente tal agente pertenceria à família dos Potyvirus, que infecta os dois gêneros, como o vírus Y da batata no gênero *Solanum* e o vírus do milho anão mosaico no gênero *Sorghum*, que poderia gerar um quadro de co-infecção (KAITISHA, 2003). Este processo pode estar ligado com a própria separação deste grupo das outras plantas e explica a sua proximidade com as monocotiledôneas.

O terceiro evento recombinante (valor de P_{SiScan} $9,2 \times 10^{-16}$) mostra a sequência de Lycopodiopsidae (*S. moellendorffii*) derivada de um segundo evento recombinante, além do 1°. Este evento envolve a sequência *V. vinifera* como parental. Entretanto, *S. moellendorffii* distingue-se de seu "parental" a um nível muito ancestral, na transição das plantas vasculares e com sementes (Spermatopsidae). Dessa forma, o evento recombinante 3 é considerado recente e poderia ser atribuído a um evento ancestral do intercruzamento entre estas espécies, ou uma recente troca mediada por sua microbiota comum, não excluindo qualquer vetor em potencial. Com efeito, somos levados a acreditar que, apesar de aparentemente não existir qualquer relação entre essas OTUs, possa haver um elevado fluxo genético.

Tabela 9. Eventos recombinatoriais nas sequências de PAL.

Evento Recombinante	BKP		Parental (% similaridade)	Valor de P (Teste)	Teste						
	Início	Fim			RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	3Seq
1	1798	1941	aor (74,5)	$2,5 \times 10^{-5}$ (BootScan)	20	12	20	20	18	-	-
2	2290	2370	sbi (94,7)	$3,1 \times 10^{-6}$ (BootScan)	2	2	1	1	1	-	-
3	1919	2188	vvi (87,8)	$9,2 \times 10^{-16}$ (SiScan)	1	-	1	1	1	1	-

* aor - *A. oryzae*, sbi - *Sorghum bicolor*, vvi - *Vitis vinifera*.

Capítulo 3. DESIGN DE UMA CONSTRUÇÃO QUIMÉRICA PARA SUBSTITUIÇÃO DE SUBSTRATO: CONVERSÃO DE UMA HAL EM PAL

MATERIAL E MÉTODOS

Design da proteína mPAL_c1

A ideia foi sumarizada na Figura 16, e resumidamente é a utilização de um *core* conservado de PAL com presumida função catalítica, como aquele previamente observado na análise da seleção códon-a-códon do capítulo 2, Figura 14, para alteração da preferência de substrato de uma HAL para uma PAL, por meio da substituição do *core* original de HAL (segmento de alta homologia) do *backbone* na enzima quimérica final.

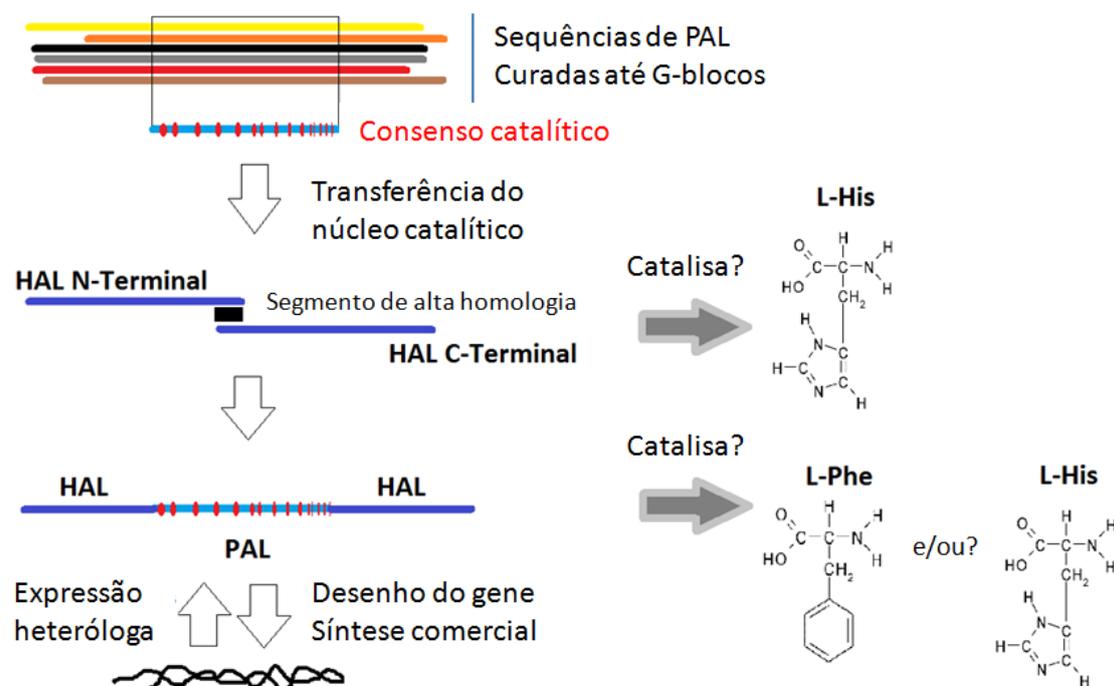


Figura 16. Resumo gráfico da construção de uma PAL híbrida. Neste esquema mostramos que a partir de um alinhamento curado de um *data set* de sequências de PAL

foi retirado um *core* catalítico, que foi "transplantado" sobre um região de homologia dentro de uma sequência de HAL. Assim, obtivemos uma enzima híbrida que possui *core* catalítico de PAL e estrutura geral de HAL. A partir dessa sequência de proteína foi gerado um gene que foi utilizado para produção heteróloga da enzima e caracterização *in vitro*.

As sequências previamente analisadas mostradas nas Tabela 4 foram utilizadas para produção da enzima quimérica. O tratamento inicial das sequências foi realizado utilizando-se a suite de aplicativos *Phylogeny.fr* (DEREEPER *et al.*, 2008) Essas sequências foram alinhadas via algoritmo MUSCLE (EDGAR, 2004) utilizando-se as seguintes penalidades para abertura de *gaps* (-2,9), extensão de *gaps* (0,0) e multiplicador de hidrofobicidade (1,2). Foi utilizado um máximo de 16 iterações, com método de agrupamento inicial para as 2 primeiras iterações por UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*), e para as demais por UPGMB (um método que combina outros dois métodos, UPGMA e *Neighbor Joining*). A distância mínima da diagonal (λ) utilizada foi de 24.

O método de UPGMA é útil para as primeiras iterações, pois assume taxas evolutivas idênticas para todas OTUs. E por meio disto, agrupa as distâncias menores e recalcula novas distâncias através de médias aritméticas para novo agrupamento. Para iterações posteriores o método conjunto de UPGMB é mais interessante, pois assume além dos critérios de UPGMA, os critérios de *Neighbor Joining*, que trabalha a partir de medidas de distâncias genéticas (SAITOU e NEI, 1987). Este método corrige o método UPGMA para seu pressuposto (frequentemente inválido) de que uma mesma taxa de evolução se aplica a todos ramos de uma filogenia. Dessa forma, a matriz de distância é ajustada para diferenças nas taxas de evolução de cada táxon (ramo) (SAITOU e IMANISHI, 1989).

Após o alinhamento, as regiões ambíguas (contendo *gaps* e/ou regiões pouco eficientemente alinhadas, por exemplo) foram removidas com *Gblocks* (v0.91b) seguindo-se os parâmetros adotados para máxima estrigência: conservação mínima das sequências para posição de flanqueamento de 85%, máximo de posições contíguas não conservadas igual a 4, comprimento mínimo dos blocos finais de 10, sem permissão de *gaps* nos blocos finais.

Os G-blocos curados foram então novamente submetidos a um alinhamento, agora pelo algoritmo de *Clustal Ω* (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), para geração de uma sequência consenso. Para tal, utilizou-se o algoritmo de *mbed-like Clustering Guide-Tree* (utiliza uma amostra das sequências de entrada e, em seguida, representa todas as sequências como vetores, permitindo a produção mais rápida de uma árvore guia), sem a destruição do alinhamento inicial. Utilizou-se 5 iterações combinadas conjuntas, com 5 iterações mediadas pela árvore-guia e outras 5 mediadas pelo algoritmo de HMM (*Hidden Markov Model*). Após o alinhamento a sequência consenso foi obtida e trimada manualmente através do BioEdit v7.1.3.0 (HALL, 1999). Em posições de ambiguidade, aminoácidos de ocorrência frequente foram utilizados, como por exemplo, em casos onde havia a ocorrência generalizada de um aminoácido de carga positiva, utilizou-se universalmente histidina (H), e em casos de ocorrência de aminoácidos de carga negativa, utilizou-se universalmente glutamato (E), da mesma forma procedeu-se com aminoácidos hidrofóbicos (utilizando-se alanina, A) e aromáticos (utilizando-se fenilalanina, F).

A sequência consenso representa uma compilação das regiões conservadas e necessárias à proteína para catálise de L-Phe. Para que a proteína fosse funcional seria necessário a adição de espaçadores peptídicos nas extremidades. A composição destes espaçadores é de fundamental importância para que a plasticidade da proteína seja respeitada bem como sua conformação final seja atingida. Assim, recorreremos a um banco de diversidade, pois como as sequências "parentais" foram obtidas de bancos de dados públicos, os espaçadores que obteríamos caso utilizássemos tais bancos poderiam ser homólogos. Assim, utilizamos espaçadores obtidos do banco de sequências metagenômicas previamente montado durante o projeto "Estudo integrado de corpos de água doce na Amazônia como visão análoga de rochas geradoras lacustres: (I) determinação por metagenoma da diversidade microbiana aquática", sob coordenação do Prof. Dr. Flávio Henrique Silva, Processo 2012/00135-0 e Termo de Cooperação 0050.0081178.13.9.

O consenso foi buscado por meio do Blastp utilizando o algoritmo DELTA-BLAST (*Domain Enhanced Lookup Time Accelerated BLAST*) contra o banco de ORFs obtidas do Lago Poraquê (03°57' S e 63°10' W). As sequências obtidas em sua maioria apresentaram similaridade com HALs, tendo em vista que a PAL é homóloga das HAL (ASANO *et al.*, 2004; ROTHER *et al.*, 2001; TAYLOR *et al.*, 1990) e que existe um consenso de que a PAL original se desenvolveu a partir de uma HAL ancestral

(FERRER *et al.*, 2008) utilizamos as sequências com maior similaridade e cobertura para a inserção.

A sequência consenso foi alinhada com os outros 2 peptídeos preditos das ORFs identificadas, por meio do algoritmo MUSCLE (EDGAR, 2004) utilizando-se os mesmos parâmetros supracitados. Após o alinhamento as regiões de espaçamento obtidas foram incorporadas à sequência final por método de *scramble*. As extensões amino-terminais e carbóxi-terminais foram mantidas sem critérios e entendidas como regiões de extensão presentes nas sequências espaçadoras e não presentes na sequência consenso (ANEXOS IV e V). A proteína final foi composta por 511 resíduos e aqui nomeada mPAL_c1.

Testes de confirmação da sequência mPAL_c1

A sequência mPAL_c1 foi buscada por Blastp contra o banco de dados de proteínas de referência (refseq_protein) e contra o banco de dados de proteínas com estrutura resolvida disponível no RCSB-PDB. As sequências "parentais" foram utilizadas para análise filogenética da mPAL_c1, por meio do software MEGA v.6.01 (TAMURA *et al.*, 2013) como descrito no capítulo 2.

A sequência mPAL_c1 ainda foi avaliada contra o banco de dados do *Protein Family Database* (<http://pfam.xfam.org/>), seguindo configurações em *default* e do servidor Interpro (<http://www.ebi.ac.uk/interpro/>). Um alinhamento detalhado dos sítios cruciais para atividade enzimática foi realizado utilizando-se os dados do PDBSum (<https://www.ebi.ac.uk/pdbsum/>) para a entrada 1W27_A, e o software Multalin (<http://multalin.toulouse.inra.fr/multalin/>).

A estrutura da mPAL_c1 foi inferida via modelagem conjunta *ab initio* e por homologia. A sequência foi carregada no servidor Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>), utilizando-se o modo intensivo. O modelo de maior confiança e menor desordem predita foi escolhido, e seu arquivo em formato ".pdb" foi utilizado. O modelo foi refinado pelo servidor KoBaMIN (<http://csb.stanford.edu/kobamin/>) que utiliza uma abordagem baseada no conhecimento de minização energética estrutural prévio por meio de *assemblies* múltiplos. Quando necessário correções manuais no arquivo de coordenadas foram realizadas, manipulações extensivas ou de menor grau de precisão foram realizadas utilizando-se o

software PyMOL v.1.3 (<http://www.pymol.org/>). Uma análise extensiva de energia resíduo-a-resíduo foi feita utilizando-se o software Swiss-PdbViewer / SPDBV (<http://spdbv.vital-it.ch/>). Os resíduos detectados como problemáticos foram então corrigidos com auxílio da ferramenta de minimização extensiva de energia livre, por auxílio da análise de anelamento de 0K a 300K com medidas a cada 0,1K, em 1000 iterações precisas. Então uma avaliação de qualidade foi realizada utilizando-se os servidores ProSA (<http://prosa.services.came.sbg.ac.at/>) e ProQ (<http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi>).

Uma análise aprofundada de comparação por alinhamento estrutural foi carreada no servidor Mistral (<http://eole2.lsce.ipsl.fr/ipht/mistral/protein.php>) por meio da comparação com as soluções estruturais 1W27_A e 3NZA_A. Os parâmetros principais utilizados durante a análise do Mistral foram: Intervalo de Interações Energéticas (2,0 Å); Tolerância do alinhamento (3,0 Å); Comprimento mais curto de segmento alinhado (100), permitindo-se eventual inversão de sequência. A solução que apresentou menor desvio energético durante o alinhamento foi utilizada para construção do tetrâmero enzimático. A estrutura foi submetida a uma dimerização simulada mantendo-se a ordem de simetria de 2:2:2, por meio do servidor SymmDock (<http://bioinfo3d.cs.tau.ac.il/SymmDock/>), que realiza um *docking* de estruturas rígidas. As 100 principais soluções foram submetidas a uma reanálise por meio do refinador SymmRef (MASHIACH-FARKASH *et al.*, 2011). As soluções foram então filtradas com base nos critérios energéticos (ACE) e de área de contato. A análise de sítios de ligação ou interação tridimensional com compostos foi realizada no 3DLigandSite (<http://www.sbg.bio.ic.ac.uk/3dligandsite/>) e ProFunc (<https://www.ebi.ac.uk/thornton-srv/databases/profunc/index.html>).

Manipulações com a sequência mPAL_c1

A sequência de aminoácidos obtida foi retrotraduzida utilizando-se o software BioEdit v7.1.3.0 (Hall, 1999). A sequência de nucleotídeos *mPAL_c1* foi otimizada para expressão conjunta em *Escherichia coli* K12 e *Picchia pastoris* com os servidores: *Rare Codon Analysis Tool* e *Optimumgene* da empresa *GenScript* (<http://www.genscript.com/cgi-bin/tools/>) e *Codon Optimization Software* da empresa IDT (<https://www.idtdna.com/CodonOpt>). O gene foi então sintetizado pela empresa GenOne (<http://genone.com.br/>) com sítios de restrição: 5' NdeI / NotI 3', e arremetido

em vetor pBlueScript II KS(+) inserido nos sítios BamHI-HindIII que foi nomeado pBS_mPAL.

O plasmídeo pBS_mPAL foi utilizado na transformação de células cálcio-competentes de *Escherichia coli* DH5 α (SAMBROOK e RUSSEL, 2001), por choque-térmico a 42°C por 90s, seguido de incubação em gelo por 2 min e recuperação em meio SOC (Tryptona 20 g/L, extrato de levedura 5 g/L, NaCl 0,5 g/L, KCl 2,5 mM, MgCl₂ 10 mM, glicose 20 mM, pH 7,0) q.s.p. 1 mL por 1h a 37°C sem agitação. As células foram plaqueadas em ágar LB (Tryptona 10 g/L, extrato de levedura 5 g/L, NaCl 10 g/L e Bacto-Ágar 15 g/L, pH 7,0) contendo ampicilina (100 μ g/mL) como agente seletivo. Após 16h a 37°C as colônias que cresceram foram recolhidas e serviram para produção de um inóculo de 5 mL de meio LB (NaCl 5g/L; triptona 5 g/L; extrato de levedura 2,5 g/L), suplementado com ampicilina como anteriormente descrito, que foi crescido a 37°C por 16h sob agitação de 250 rpm; e foi utilizado para extração de pBS_mPALc1 com o QIAquick Miniprep Kit (Qiagen).

O vetor purificado foi digerido com as enzimas NdeI e NotI em reação contendo: 1 μ g do vetor pBS_mPALc1, 5U de NdeI (Thermo), 5U de NotI (Thermo), Tampão *Orange* (Thermo) 2X e água q.s.p. 50 μ L, a 37°C por 16h. Após este período o fragmento de 1544 pb foi separado por eletroforese em gel de agarose 0,8% e purificado com auxílio do QIAquick Gel Extraction Kit (Qiagen). A reação de ligação foi realizada com 100 ng de pET28a linearizado com as mesmas enzimas, 88 ng de inserto, Tampão Ligase 1X (Invitrogen) e 1U da enzima T4 DNA ligase (Invitrogen), em volume final de 30 μ L e incubou-se a 4°C por 16h. Após este período o produto foi utilizado para transformação de células de *E. coli* DH5 α cálcio-competentes como anteriormente descrito (SAMBROOK e RUSSEL, 2001), e as colônias transformantes foram selecionadas em meio ágar LB contendo canamicina (50 μ g/mL). As colônias recombinantes foram selecionadas por meio de PCR confirmatório utilizando os primers do vetor, conforme segue: 100 ng de DNA *template*, 0,2 mM de dNTPs, 5 pmol do primer P-pT7 (5'-TAATACGACTCACTATAGGG-3'), 5 pmol do primer T-pT7 (5'-GCTAGTTATTGCTCAGCGG-3'), 1X tampão de Taq-DNA pol. (Sinapse), 1U Taq-DNA polimerase (Sinapse), água nanopura estéril q.s.p. 25 μ L. A ciclagem consistiu de desnaturação inicial a 95°C por 3 min; seguida de 35 ciclos de desnaturação a 94°C por 30s, anelamento a 50°C por 30s, extensão a 72°C por 1,5 min; e finalmente uma extensão final a 72°C por 7 min. Os amplicons foram resolvidos em gel de agarose 0,8% a 100V por 30 min, corado com brometo de etídeo 0,1 μ g/mL. O gel foi

fotografado em sistema de fotodocumentação ChemiDoc XRS System (Biorad). Os amplicons de ~1790pb sinalizaram uma inserção correta e provável no vetor final de expressão, esses clones foram utilizados para produção de mais vetor conforme anteriormente descrito, para fins de verificação.

Os vetores recombinantes então foram sequenciados utilizando-se 5 pmol dos primers P-pT7 ou T-pT7, com o kit DYEnamic ET Dye Terminator (GE Healthcare) em plataforma MEGA BACE (GE Healthcare) conforme instruções do fabricante. Após sequenciamento, o eletroforetograma foi analisado em software BioEdit v7.1.3.0 (HALL, 1999), e as bases de confiabilidade reduzida detectadas por meio dos algoritmos Phred/Phrap/Consed v.28.0 (GORDON e GREEN, 2013; GORDON *et al.*, 1998; GORDON *et al.*, 2001; EWING *et al.*, 1998; EWING e GREEN, 1998) foram trimadas. A sequência foi então inserida no programa VecScreen (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/>) para eliminação de sequências do vetor, e então foram alinhadas no software MEGA v.6.01 (TAMURA *et al.*, 2013) para produção da sequência flanqueadora confirmada (no mínimo 500pb iniciais e finais). O vetor final de expressão pmPAL (vetor pET28a contendo fragmento *mPALc1*) de 6,84 kpb foi utilizado nas próximas manipulações.

Expressão recombinante de mPAL_c1

O plasmídeo pmPAL foi utilizado para transformar por meio de choque-térmico células cálcio-competentes de *E. coli* Rosetta(DE3) (SAMBROOK e RUSSEL, 2001). O meio de seleção para essas células foi composto por ágar LB suplementado com canamicina (50 µg/mL) e cloranfenicol (34 µg/mL).

A indução foi efetuada utilizando-se um pré-inóculo crescido *overnight* a 37°C/250 rpm em meio LB (NaCl 5g/L; triptona 5 g/L; extrato de levedura 2.5 g/L) suplementado com cloranfenicol e canamicina (conforme anteriormente referido), seguido de uma diluição de 1:10 em 500 mL de meio LB fresco contendo os mesmos antibióticos. O crescimento celular foi monitorado a 600 nm em um espectrofotômetro UV-Vis (Hitachi) até que a densidade óptica de 0.5 fosse atingida. Em seguida, foi adicionado isopropil β-D-1-tiogalactopiranosídeo (IPTG; Sigma) a uma concentração final de 0.4 mM. O meio foi novamente incubado a 37°C/200 rpm durante o tempo ótimo de produção previamente determinado como 3h.

Após o período de indução as células foram pelletizadas por meio de centrifugação a 12.000xg por 10 min a 4°C em centrífuga Sorvall, rotor SS-34. O pellet foi ressuscitado em gelo cuidadosamente em 1:100 ($v_f:v_i$) de Tampão de Lise (Tris-HCl 10 mM, NaH₂PO₄ 50 mM, NaCl 100 mM, pH 8). A suspensão foi então sonicada em sonicador Sonic Dismembrator 500 (Fisher Scientific) no gelo por 6 min, em pulsos de 59,9s e intervalos de 30,0s (Amplitude de 20%). Após este tratamento, as células foram centrifugadas a 20.200xg por 20 min e o sobrenadante foi filtrado a 0,45 µm (Millipore). A amostra de sobrenadante foi considerada como a porção de proteína solúvel e a porção do *pellet* considerada composta por proteínas insolúveis.

A porção insolúvel foi lavada duas vezes com 25 mL de tampão de lise, com centrifugações nas mesmas condições. Após as lavagens, o *pellet* foi ressuscitado com tampão de lise em um volume final de 25 mL, a esta suspensão, adicionou-se ureia em pérolas para uma concentração final de 8M num volume final de 50 mL (atingido por meio de adição de tampão de lise estéril). A mistura foi agitada cuidadosamente à temperatura ambiente por 45 min e a etapa de centrifugação foi repetida. O sobrenadante foi filtrado a 0,45 µm (Millipore), e foi considerada como a porção de proteína desnaturada solúvel.

A porção de proteína desnaturada solúvel foi utilizada para o *refolding*. Para isso, inicialmente realizamos a purificação da proteína por cromatografia de afinidade à metal imobilizado (IMAC) utilizando 3 mL de resina Ni²⁺-NTA *superflow* (Qiagen) seguindo-se as recomendações do "*The Qiaexpressionist Handbook Fifth Edition* (06/2003)" quanto à purificação em ambiente desnaturante por meio de gradiente de pH. Após isto, as frações purificadas foram reunidas e adotamos 3 métodos diferentes:

1. Ultrafiltração utilizando colunas Vivaspin (MWCO 10 kDa) até uma concentração de 10 vezes a partir de centrifugação a 3.260xg a 4°C;
2. Diálise em tampão de lise utilizando um gradiente de uréia (7M, 6,5M, 4,0M, 2,0M, 0,5M, 0M) em diálises em membrana Snake-Skin (Pierce, MWCO 3kDa) de 3h a 4°C sob agitação cuidadosa (caso houvesse a precipitação da proteína a diálise seria parada e uma centrifugação de alta velocidade a 20.200xg a 4°C durante 25 min e o sobrenadante filtrado a 0,45 µm);

3. Diluição rápida: A porção solúvel foi diluída numa proporção de 1:50 ($v_i:v_f$) em tampão de lise esterilizado, por gotejamento (1 mL/min), sob agitação cuidadosa. A diluição final foi filtrada a 0,45 μM , e mantida *overnight* a 4°C. A proteína mPAL_c1 proteína foi purificada novamente em uma nova coluna contendo 3 mL de resina Ni²⁺-NTA *superflow* (Qiagen), previamente equilibrada com tampão de lise. Após recolher o eluato, realizou-se um passo de lavagem com dois volumes de tampão de lise contendo 25 mM de imidazol (frações 25A-B). A eluição da proteína foi efetuada com 4 volumes de tampão de lise contendo 300 mM de imidazol (frações 300A-D). A proteína purificada foi dialisada em membrana Snake-Skin (Pierce, MWCO 3kDa) durante a noite contra 170x de Tampão de diálise (NaH₂PO₄ 25 mM, NaCl 50 mM, pH 8.0) a 4°C sob agitação suave, e este processo foi repetido durante 3 h nas mesmas condições. O processo de concentração de proteínas foi feito em filtro Vivaspin 200 (GE Healthcare, MWCO 10 kDa) a 4°C até concentrações finais próximas a 5 μM serem atingidas. Após o processo, o material foi filtrado a 0.22 μm (Millipore) para esterilização e foi acondicionado em gelo até o uso. Um esquema final do processo de purificação é mostrado na Figura 17.

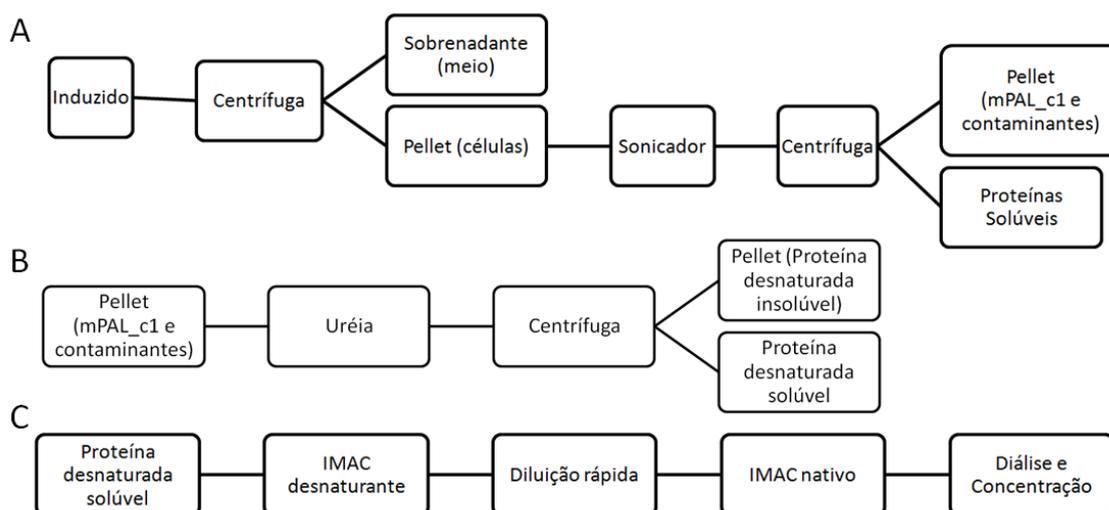


Figura 17. Esquema de purificação da mPAL_c1 por meio de *refolding* por diluição rápida. Inicialmente as células são concentradas e lisadas por ultrassom (A), depois a porção insolúvel é tratada com uréia (B) e a porção solúvel deste tratamento é submetida a uma IMAC desnaturante para concentração da proteína, então o produto é renovelado por meio de diluição rápida e passa por uma IMAC nativa para

concentração da mPAL_c1 pura (C). A enzima por fim é dialisada contra tampão fosfato e concentrada.

Os SDS-PAGEs (12% e 15%) e Western-Blot (WB) foram realizados conforme descrito por Sambrook e Russel (2001). Brevemente, o procedimento do WB consistiu na transferência das proteínas para uma membrana de PVDF (Pierce Biotechnology), previamente ativada com metanol, via eletroforese vertical a 115 mA durante 4h e as bandas de interesse foram visualizadas por pré-coloração de Ponceau 5%. Após descoloração e bloqueio com leite desnatado (Molico), as bandas foram então marcadas com anti-corpo anti-His-Tag de camundongo (dil. 1:2500; Thermo-Fisher) e para revelação utilizou-se o *kit Amersham ECL Western Blotting System* (GE Healthcare) com o anti-corpo secundário anti-mouse-HRP numa diluição de 1:10.000. Os sistemas de detecção foram fotografados e documentados pelo sistema de fotodocumentação *ChemiDoc XRS System* (Biorad).

Determinação da concentração de mPAL_c1 purificada

A concentração da proteína recombinante foi determinada com base na absorção a 280nm, medido em um espectrofotômetro U-2001 Hitachi, utilizando o coeficiente de extinção molar teórico (ϵ) calculado a partir da composição de aminoácidos da proteína por meio da equação de Pace *et al.* (1995). O coeficiente de extinção molar teórico da proteína no estado nativo para mPAL_c1 foi $\epsilon_{280} = 30367,69 \text{ M}^{-1}\text{cm}^{-1}$ (assumindo-se que todos os pares de cisteína formem cistinas, assim Abs 1 mg/mL seria de 0,5219). Para efeitos de cálculo a massa da enzima foi estimada como 58,186 kDa. Assim, a concentração mássica da enzima é dada pela fórmula derivada da lei de Beer-Lambert:

$$[\mu\text{g}/\mu\text{L}] = (\text{MW} * \text{Abs}_{280\text{nm}}) / (\epsilon_{280} * l) = 19,161 * \text{Abs}_{280\text{nm}}$$

Onde: a concentração em $\mu\text{g}/\mu\text{L}$ é dada pela multiplicação da massa molecular (MW) em Da multiplicado pela absorvância a 280 nm, dividido por uma multiplicação do coeficiente de extinção molar versus o caminho óptico em cm (0,1 cm em aparelho Nanodrop ND-1000).

Medição de estado oligomérico

Diferentes populações oligoméricas de uma proteína podem ser separadas pela cromatografia de exclusão molecular, cujo princípio se estabelece na porosidade do polímero componente da matriz da fase estacionária que retarda moléculas de acordo com seu raio hidrodinâmico, diretamente ligado à forma e tamanho molecular (ERICKSON, 2009).

Foi utilizada uma coluna de gel filtração HiPrep Sephacryl S300 HR 16/60 (GE Healthcare) a fim de se determinar o estado de oligomerização da mPAL_c1 renaturada. A coluna foi previamente equilibrada com o tampão de diálise. A eluição da amostra foi isocrática com 1,5 volume de coluna sob uma taxa de fluxo de 0,5 mL/min de tampão de diálise (NaH₂PO₄ 25 mM, NaCl 50 mM, pH 8,0) e o volume de amostra injetado de 1.500 µL.

Para calcular o raio hidrodinâmico (Rh) a coluna foi calibrada com os padrões protéicos (Sigma-Aldrich) para coluna de gel filtração dispostos na Tabela 2. Como recomendado pelo manual do fabricante, os padrões foram preparados em uma mistura com concentração final de 0,3 mg/mL de proteínas. O composto *Blue Dextran* 2000 (GE Healthcare) foi empregado para a determinação do volume "morto" da coluna (V_o). Sabendo-se os volumes de eluição de cada proteína (V_e), foi possível calcular os valores do coeficiente de partição (K_{av}), a partir da fórmula $K_{av} = (V_e - V_o) / (V_T - V_o)$, sendo V_T o volume total da coluna. Os valores de K_{av} foram utilizados para a construção da curva de calibração, K_{av} versus log(Rh) (ANEXO III), permitindo o cálculo do raio hidrodinâmico e estimação da massa molecular aparente das proteínas de interesse por regressão linear.

Tabela 10. Principais proteínas padrão utilizadas no experimento de calibração.

Proteína	MW (kDa)	Rh (nm)	K _{av}
Miosina	205	12,0	-
B-galactosidase	116	6,9	0,32
Fosforilase B	97,4	6,0	0,42
BSA	66	3,6	0,53
Ovoalbumina	45	2,9	0,79
Anidrase Carbônica	29	2,0	1,04

Medição da restituição do co-fator MIO por varredura espectroscópica

A restituição do cofator MIO (4-metilideno imidazol-5-ona) foi observada conforme descrito por Kovács *et al.* (2014) e Poppe (2001), com ligeiras modificações. O sinal foi detectado por espectros de varredura UV-Vis (220 nm a 340 nm) coletados num espectrofotômetro UV-Vis (Hitachi) a partir da diferencial de curvas entre amostras obtidas por renaturação por meio de diluição rápida (Rd) ou por meio de diálise e amostras desnaturadas (em uréia 8M). As curvas foram confirmadas em Nanodrop ND1000 (Thermo).

Medição do $\lambda_{Máx}$ da emissão de triptofanos

Os ensaios de emissão do triptofano foram realizados a fim de se observar a posição relativa dos resíduos de triptofano na proteína, podendo assim nos revelar dados sobre o seu enovelamento. Os ensaios foram realizados em um espectrofluorímetro F-2500 (Hitachi), sob comprimento de onda de excitação de 295 nm e de varredura da emissão variando entre 300 nm a 400 nm, a 20°C. Para isso, utilizou-se uma cubeta de quartzo com caminho óptico de 10 mm, fendas de excitação e de emissão de 2,5 nm, uma velocidade de varredura de 60 nm/min, e tempo de resposta de 0,08 s. A tensão PMT utilizada foi de 400V. Todas as amostras (desnaturada) e renaturada *in vitro* foram medidas sempre a 5,8 μ M de mPAL_c1 em tampão de diálise (NaH₂PO₄ 25 mM, NaCl 50 mM, pH 8,0).

Dicroísmo circular

As medições em Dicroísmo Circular (CD) foram realizadas empregando a concentração de 2,6 μ M de proteína recombinante em tampão de diálise (NaH₂PO₄ 25 mM, NaCl 50 mM, pH 8,0) para observar as diferenças estruturais induzidas pela concentração de proteína. Os espectros de CD foram monitorados com um espectropolarímetro Jasco (J-815) a 20°C com resolução de 0,2 nm, e cubetas com caminho óptico de 0,2 cm de comprimento. Um total de 40 acumulações para cada espectro UV entre 194 e 260 nm foi recolhido, determinando assim a estrutura secundária da proteína após uma resultante da duplicata e subtração da curva obtida

somente com o tampão. A voltagem do amplificador óptico foi monitorada e nunca ultrapassou 700V (HT Volt. < 700V). O software CDNN, v.2.1 (BOHM, 1997) foi utilizado para conversão das medidas e deconvolução das curvas. Para a análise utilizou-se a biblioteca de espectros base "*Net-using 33 base spectra*" para dados complexos de CD, com um *Jack-Knifing* de casos máximos para teste de 1 e frequência de aumento de varredura de 10. Os dados de espectros de CD foram relatados como elipsidade residual média $[\theta]_{MRW(\lambda)}$, em grau.cm²/dmol, que foram calculados a partir de valores da elipsidade em milligraus (θ), utilizando a fórmula Kelly *et al.* (2005):

$$[\theta]_{MRW(\lambda)} = \theta \cdot MW / (NRes \cdot c \cdot l)$$

Onde: $[\theta]$: elipsidade molar; MW: peso molecular em Daltons; l : caminho óptico da cubeta; NRes: número de aminoácidos; c concentração da proteína em mg/ml).

Medição de atividade da mPAL_c1 purificada

O ensaio para determinar a atividade enzimática da mPAL_c1 foi feito de acordo com o método de Zimmerman e Hahlbrock (1975) que depende das diferenças espectrais entre o substrato e o produto, com as modificações posteriores de Kóvacs *et al.* (2014) e nossas modificações. A absorção de ácido trans-cinâmico e ácido urocânico a 270 e 290 nm é tipicamente maior do que a de L-Phe e L-His, respectivamente. Portanto, a reação enzimática foi acompanhada pelo Nanodrop ND-1000 (Thermo) por meio da função espectrofotômetro UV-Vis, em ambos os comprimentos de onda. O progresso da eliminação de amônia a partir de L-Phe foi monitorado por detecção da produção de t-CA ($\epsilon_{270nm} = 19.730 \text{ M}^{-1}\text{cm}^{-1}$; $\epsilon_{290nm} = 10^4 \text{ M}^{-1}\text{cm}^{-1}$ a 25°C). O progresso da eliminação de amônia a partir de L-His foi monitorado por detecção da produção de ácido urocânico ($\epsilon_{277nm} = 13.600 \text{ M}^{-1}\text{cm}^{-1}$ a 25°C). A reação procedeu-se em microtubos protegidos da luz. Utilizou-se cerca de 100-300ng de mPAL_c1 purificada e foi adicionada a um volume final de reação final de 25 μL com tampão de Davies (ácido cítrico 27 mM, tetraborato de sódio 9 mM, Tris 25 mM, KCl 25 mM - DAVIES, 1959) em diferentes pHs, contendo diferentes concentrações de L-Phe (0-5 mM) ou L-His (0-5mM); e diferentes concentrações de MnCl₂ (0-6 mM). O aumento de absorção foi

registrada após 30 min para reações contendo L-Phe e depois de 90 min para reações contendo L-His. As medições foram realizadas à temperatura ambiente (25°C). As reações foram paradas por meio da adição de HCl 4M até concentração final na reação de 1M, seguido por uma centrifugação a 16.000xg a 4°C por 30 min e acondicionamento no gelo até a leitura.

Para confirmação de atividade utilizou-se um sistema de duplo branco: inicialmente o equipamento foi zerado com tampão de reação, após isto adotamos a medida em triplicata de reações com e sem a enzima, paradas no tempo zero e no tempo final de medida. Assim, obtivemos a atividade pela equação:

$$A_{(U)} = [(Enz_{.t} + S_{t=0}) - (Enz_{.t=0} + S_t)] * \kappa, \text{ onde:}$$

"A" refere-se à Atividade medida em unidades internacionais; $Enz_{.t}$ e $Enz_{.t=0}$ referem-se respectivamente à absorção da reação contendo enzima parada nos tempos final e inicial; S_t e $S_{t=0}$ referem-se respectivamente à absorção da reação sem a enzima parada nos tempos final e inicial; κ refere-se à constante de reação que depende das condições de reação, diferentes constantes são mostradas na Tabela 11.

Tabela 11. Constantes de reação para cálculo da atividade em diferentes condições.

As constantes mostradas foram calculadas para um volume de reação de 25 μ L e os tempos de reação para L-Phe e L-His de 30 e 90 min, respectivamente.

Substrato	λ / nm	κ / μmol.min⁻¹
L-Phe	270	4,22x10 ⁻⁴
L-Phe	290	8,33x10 ⁻⁴
L-His	277	2,04x10 ⁻⁴

Assim defimos aqui como unidade internacional (U) de mPAL_c1, como a quantidade de enzima necessária para desaminar 1,0 μ mol de L-Phe/L-His a trans-cinamato/urocanato e amônia por minuto. Quando conveniente, a atividade da enzima foi expressa como atividade relativa.

As comparações entre a enzima comercial de *Rhodotorula glutinis* (Sigma-Aldrich) foram realizadas nas mesmas condições ótimas para reação com a mPAL_c1. Todas as reações foram feitas em triplicata, e o tratamento estatístico foi realizado utilizando-se o GraphPad Prisma v.5.0.

Efeitos da temperatura, pH e inibição por amônia

O teste do tempo ótimo de reação foi realizado num pH 7,5 com tampão de Davies (DAVIES, 1959) a 30°C contendo 300 ng de enzima e 2,5 mM de substrato (L-Phe ou L-His). A variação de absorvância foi verificada de 15 em 15 min até 3 h.

Todos os testes desta seção foram realizados em triplicata no mínimo. A determinação do pH ótimo a 30°C foi realizada para ambos os substratos utilizando-se condições similares as da determinação de tempo ótimo, variando o pH do tampão de Davies (DAVIES, 1959) sob um *ranging* de 3,0 a 12,0. A temperatura ótima foi determinada em reações contendo as mesmas quantidades de enzima e substrato de interesse anteriormente citadas, e foram realizadas sob o pH previamente definido como ótimo para cada substrato, sob diferentes temperaturas (30°C a 80°C), em termociclador T100 (Biorad). Após determinação de pH ótimo e temperatura ótima, todas as reações foram realizadas em condições ideais para seu respectivo substrato.

A confirmação do efeito inibidor exercido pela amônia, foi verificado pela adição de quantidades determinadas de NH_4Cl à reação contendo 3,5 mM de L-Phe e de L-His como respectivos substratos, medindo-se a atividade relativa com base na atividade na ausência deste composto, sob condições ideais.

Determinação das constantes cinéticas da enzima (K_M , $V_{M\acute{a}x}$. e K_{cat})

Os valores de $V_{M\acute{a}x}$, K_M e K_{cat} foram obtidos a partir dos testes de bondade de ajuste de curvas realizado no software Graph Pad Prism 5.0, por meio do uso de dinâmica de atividade enzimática Michaeliana, linear e não-linear. Tais dados foram obtidos por meio da medida de atividade enzimática após variação das concentrações de L-Phe e L-His sob condições ótimas de pH e temperatura.

RESULTADOS E DISCUSSÃO

Análise bioinformática

Utilizando síntese gênica após uma abordagem de *design in silico*, as sequências espaçadoras de HAL foram fusionadas ao *core* catalítico de PAL previamente isolado (Figura 18; ANEXOS IV e V). A proteína final mPAL_c1 possui 532 resíduos (incluindo a *His-Tag* amino-terminal do vetor pET28a), e o gene teve comprimento completo de 1550 pb. As sequências foram depositadas no *European Nucleotide Archive*: mPAL_c1 (LT159967; CVO82816.1) e as ORFs parentais de HAL (LT159968; LT159969). Os principais parâmetros da mPAL_c1 foram MW de 58,187 kDa e pI de 5,71. O Blastp do NCBI revela homologia considerável com PALs vegetais, sem resultados envolvendo HALs, e sua principal identidade foi com a sequência de PAL de *Musa acuminata* (E-valor de 0e+00, identidade de 72% e cobertura de 80%). Apesar da alta similaridade que a mPAL_c1 apresenta com outras PAL de eucariotos superiores, seu comprimento foi reduzido cerca de 24,43% da média observada no *data set* original (Tabela 4).

Muitos sítios essenciais para atividade de PAL (Tabela 12) foram conservados no processo do *design*, e são mais similares com as sequências de plantas nos sítios de ciclização do cofator MIO (III), e reconhecimento de L-Phe. Nos sítios de ciclização do cofator MIO (I) e (II) há uma fusão nos resíduos em que a identidade entre plantas e fungos é evidente. O sítio de acidificação COO⁻ foi considerado misto.

Assim, o sítio ativo da mPAL_c1 foi determinado como H75, N187, Q284, Y289, R294, F348 e Q434. Os precursores do cofator MIO, foram A135, S136, G137. Estes resíduos do sítio ativo são importantes para ligação do substrato e catálise (POPPE, 2001; GOLDSOON-BARNABY e SCAMAN, 2013).

O sítio MIO (I) compõe um pocket hidrofóbico de seleção do substrato bastante específico, onde a sequência indica atividade enzimática presente, no caso de PAL é a sequência FL, a sequência HQ indica atividade de PTAL e HL atividade de TAL (GOLDSOON-BARNABY e SCAMAN, 2013), entretanto, na mPAL_c1 tal sítio é misto (SH), e sua afinidade deverá ser testada no futuro para substratos alternativos, tais como, L-Tyr e outros aminoácidos. Os resultados multi-alinhamento da mPAL_c1 com outras PAL e HAL de estrutura resolvida são mostrados no ANEXO VI.

```

      1      10      20      30      40      50
8LP_44363.118104_827 AANKITACRNYLDKKLDDSAVFGINTGFGFLQNVQIDATQTEQLQYNLLMSH
Core_PAL
8LP_5286.33660_929 .....

      60      70      80      90      100
8LP_44363.118104_827 ACGLGQVPEVEIVRLMLLLKIKSLSYGNISGVQLQTVTRLIELFNHDVFPPIIFTQ
Core_PAL
8LP_5286.33660_929 .....PHSATRAAMLVRIINTLLQGYSGIRFEILEAITKLLNHNITPCLPLR

110      120      130      140      150      160
8LP_44363.118104_827 GSLGASGDLAPLSHLSLPLIGLGEVIHHQKKQLAKHVLQHFNWQPISLKSKEGL
Core_PAL
8LP_5286.33660_929 GTISASGDLVPLSYIAGLLTGF.....ELQPKLKEGL

      170      180      190      200      210
8LP_44363.118104_827 ALINGTQFM SAYGMYC L YKASH L I K M A D Y I S A L S F D A F D C S T E A L H P L I H K I R P
Core_PAL
8LP_5286.33660_929 ALVNGTAVGSLASLVLFEEANILAVLSEVLSAIFAEVMQGKPEFTDPLTHKLRH

      220      230      240      250      260
8LP_44363.118104_827 HKGVNTAKRNILQYLRSEI...SQKKKQIQDPYSFRCPQVHGAITKDTFEYV
Core_PAL
8LP_5286.33660_929 HPQOIEAAATMEHILDGS.....QDRYALRTSPQWLGPIEIVIRFA
QRGQIDVAAYRDLITDSEIAHSHEDCSKVQDEYSLRCQPVVMGACLTIQIRA

270
8LP_44363.118104_827 LNVFEREIN.....GNNFQGTPIGVSMDNTRLALASIGKLMF
Core_PAL
8LP_5286.33660_929 AEVLEIEANAVSDNPLVFAEQGDVISTGNNFHAEPVAMAA DNLA LA I A I G S L S E

8LP_44363.118104_827 .....
Core_PAL
8LP_5286.33660_929 A Q F S E L V N D F Y N N G L P S N L N P S L D Y G F K G A E I A M A S Y C S E L Q Y L A N P V T S H V Q
R R I S M M D R H M S Q L P P F L V A N G G V N S G F M I A Q V T A A A L A S D N K A L A H P A S V D S L

8LP_44363.118104_827 .....
Core_PAL
8LP_5286.33660_929 S A E Q H N Q D V N S L G L I S S R K T A E A V D I L K L M S A T F L Y A L C Q A I D L R G . . E E I D K V
P T S A N Q E D H V S M A P N A G K R L W Y M A D N V R G I L A V E W L G A C Q G L D F R E G L K S S P K L

8LP_44363.118104_827 .....
Core_PAL
8LP_5286.33660_929 E Q A R K I L R A Q V P Y Y S E D R F F A P D I E Q A S E L L S S G C L N E L I I P K L L P S L S E V

```

Figura 18. Montagem da mPAL_c1 mostrando regiões onde o core conservado foi inserido entre as ORFs da HAL que o recebeu. O core consenso com posições predominantes positivas substituídas por His foram alinhadas com as ORFs amino (8LP_44363.118104) e carbóxi terminais (8LP_5266.33660) da HAL que serviu como backbone obtida do metagenoma do Lago Poraquê. A partir deste alinhamento as zonas de alta identidade foram utilizadas como regiões para transplante do core de PAL.

Tabela 12. Comparação dos resíduos em torno dos sítios ativos da PAL. Resíduos catalíticos são mostrados em negrito.

Espécie	MIO	COO⁻	L-Phe			
<i>O. sativa</i>	RFLNAG	TVTASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>O. sativa</i>	RFLNAG	TITASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>S. bicolor</i>	RFLNAG	TVTASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>S. italica</i>	RFLNAG	TVTASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>Z. mays</i>	RFLNAG	TVTASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>B. distachyon</i>	RFLNAG	TITASGDLV	MVNGT	LMKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>R. communis</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>P. trichocarpa</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>G. max</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>C. arietinum</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>M. truncatula</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>C. sativus</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>A. thaliana</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>C. clementina</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>C. sinensis</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>V. vinifera</i>	RFLNAG	TVTASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>S. lycopersicum</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>S. tuberosum</i>	RFLNAG	TITASGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>F. vesca</i>	RFLNAG	TISSGDLV	LVNGT	LQKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>S. moellendorffii</i>	RFLNAG	TITASGDLV	IVNGT	LKKPKQDRY--ALR--TSPQ	GGNFQGT	HNQDV
<i>N. punctiforme</i>	WFLKSG	SIGASGDLV	MMNGT	GKDLIQDRY--SLR--CLAQ	GGNFLGQ	FNQNI
<i>A. oryzae</i>	QHHNAA	SISASGDLT	LLNGT	HGGLAQDRY--ALR--TSTQ	GGNFQAA	HNQGL
<i>A. niger</i>	QHHNAG	SISASGDLT	LLNGT	EVGLAQDRY--AIR--TTPQ	GGNYQAA	HNQSI
<i>A. fumigatus</i>	QLLQAG	SISASGDLM	LVNGT	RPGLIQDRY--ALR--TVPQ	GGNFQAV	HNQSV
<i>N. fischeri</i>	-----	SISASGDLM	LVNGT	RPGLIQDRY--ALR--TVPQ	GGNFQAV	HNQSV
<i>A. clavatus</i>	QLLQAG	SISASGLI	LVNGT	RAGLIQDRY--AVR--TVPQ	GGNFQAV	HNQAV
<i>U. reesii</i>	QLLQSG	TISASGDLM	LVNGT	RQGLVQDRY--ALR--TASQ	GGNFQAT	HNQAV
<i>V. albo-atrum</i>	QHTQSA	TISASGDLM	LINGT	GEGLWQDRY--STR--TAPQ	GGNFQAT	HNQGI
<i>C. cinerea</i>	QHQHVG	SISASGDLS	ILNGT	KYNLRQDRY--PLR--TSPQ	GGNFQAM	HNQAV
<i>U. maydis</i>	QMQQCG	SISASGDLS	LLNGT	NGTLRQDRY--TLR--TASQ	GGNFQAM	ANQAI
<i>M. oryzae</i>	SHLTCG	SISASGDLS	IVNGT	GFTLRQDRY--SIR--TASQ	GGNFQAR	GNQAI
<i>P. graminis</i>	EHLLAG	TISASGDLS	IANGT	KQILRQDRY--PLR--TSPQ	GGNFQAT	HNQSI
mPAL_c1	MSHACG	TISASGDLV	LVNGT	DGSQDNPRYNASLRNCTSPQ	GGNFQGT	HNQDV

A fim de compreender melhor a proteína mPAL_c1, uma modelagem por homologia foi realizada (Figura 19). O modelo teve de passar por um processo de otimização energética e conformacional (ANEXO VII), sendo que sua conformação final apresentou um *Z-score* de -7,13, *LG-Score* de 2,119 e *MaxSub* de 0,104. Esse conjunto de valores mostra que o modelo utilizado possui ainda imperfeições, entretanto, ele é muito próximo de modelos obtidos experimentalmente via difração de raios-X, como observado pelo algoritmo do PROSA (ANEXO VII).

Quanto à forma, a mPAL_c1 se revela uma proteína oblongada, cujas dimensões são para o: monômero (12,7x4,05x4,64 nm), dímero (15,23x6,36x5,01 nm) e tetramero (15,51x6,01x9,58 nm). O modelo construído foi capaz de cobrir a mPAL_c1 dos resíduos 21-531 e foi composto por: α -hélices (54%), *turns* (32%), *random coils* (11%), folhas β (2%) e pontes β (1%).

A estrutura possui zonas de irregularidade que não puderam ser corrigidas entre os resíduos 15-30 e 100-108 que formam *random coils* distribuídos sob a zona de projeção amino terminal que compreende os resíduos 1 - 193 (Figura 20), relativamente comum em organismos procariotos (KOVÁCS *et al.*, 2014). A extensão carbóxi terminal típica das PAL de eucariotos superiores (RITTER e SCHULZ, 2004), teoricamente compreendida entre os resíduos 474 - 532 da mPAL_c1, forma um arco sobre o sítio ativo e acredita-se que funcione como uma blindagem para restrição da entrada e saída de produtos e substrato (RITTER e SCHULZ, 2004). Entretanto, ainda há a hipótese proposta por Pilbák *et al.* (2006) de que esta extensão C-terminal possa diminuir o tempo de vida da PAL, desestabilizando a conformação conservada da Y110 (PAL de *Petroselinum crispum*) ou da Y45 da mPAL_c1.

Numa análise utilizando o PDBSum, 3DLigandSite e ProFunc que analisam a compatibilidade estrutural de *pockets* superficiais da estrutura com um banco de estruturas de possíveis ligandos, conseguimos determinar os resíduos que formam as cavidades do sítio ativo da enzima e de ligação à substrato que são mostrados na Tabela 13. Alguns destes resíduos já haviam sido detectados pela análise da sequência linear feita anteriormente e confirmam as teorias de Poppe (2001) e outros (RITTER e SCHULZ, 2004; WANG *et al.*, 2008; MOFFITT *et al.*, 2007). Além dos sítios convencionais de ligação à L-Phe que foram consistentes com o sítio encontrado por Wang e colaboradores (2008), obtivemos dois sítios adicionais de ligação ao t-CA, o que sustenta a hipótese de que possa haver inibição da PAL por esse composto,

representando um mecanismo de *feedback* (MACDONALD e CUNHA, 2007) por meio da mudança estrutural induzível por produto.

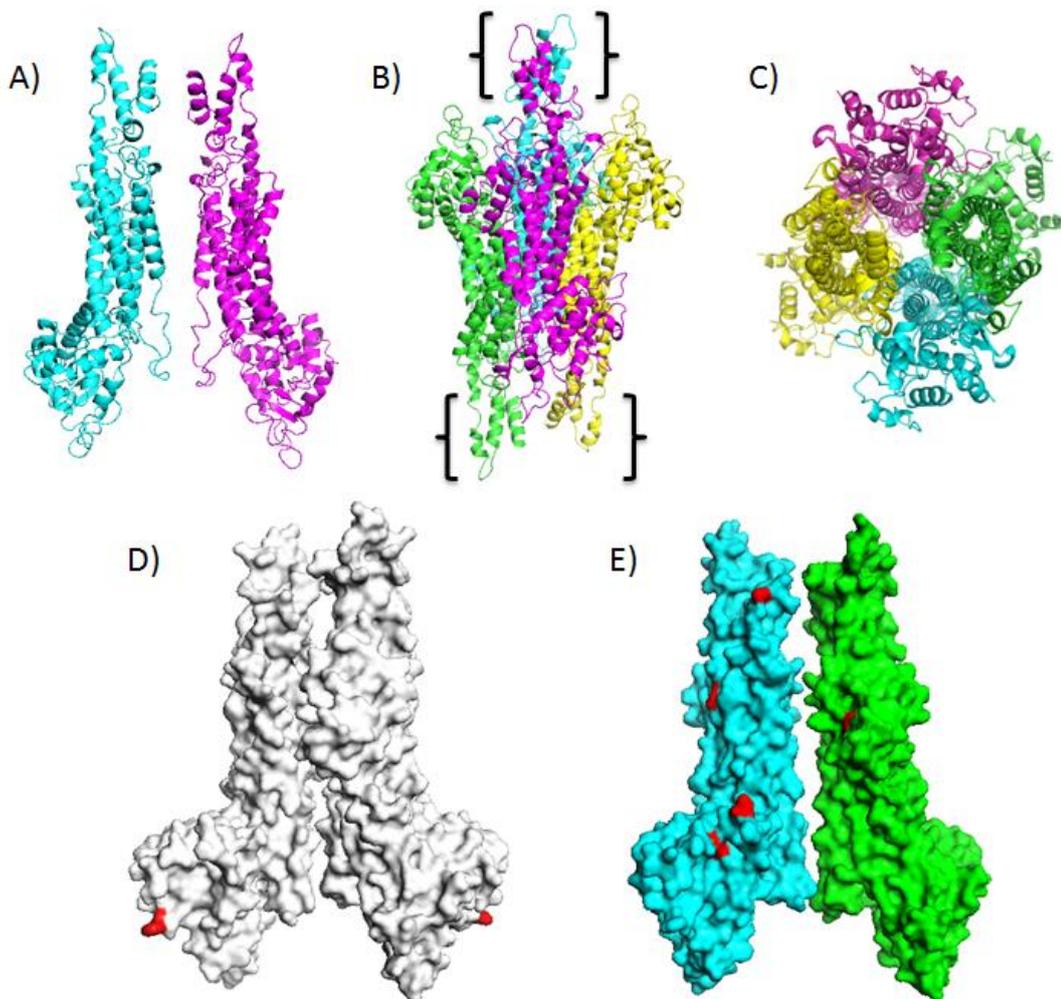


Figura 19. Estrutura mPAL_c1. São mostrados o homodímero (A) e homotetrâmero com as extensões carbóxi e amino-terminais ressaltadas pelas chaves em preto (B). O arranjo ortogonal dos monômeros na formação do homotetrâmero (C) mostram a simetria 2:2:2. Os triptofanos são evidenciados em vermelho (D), entretanto apenas o W172 está exposto na superfície proteica (são mostrados dímeros alinhados à estrutura do tetrâmero). As cisteínas são mostradas em dímeros alinhados à estrutura do tetrâmero em vermelho (E) e evidenciam que apenas três (C280, C466 e C412) estão expostas na estrutura final.

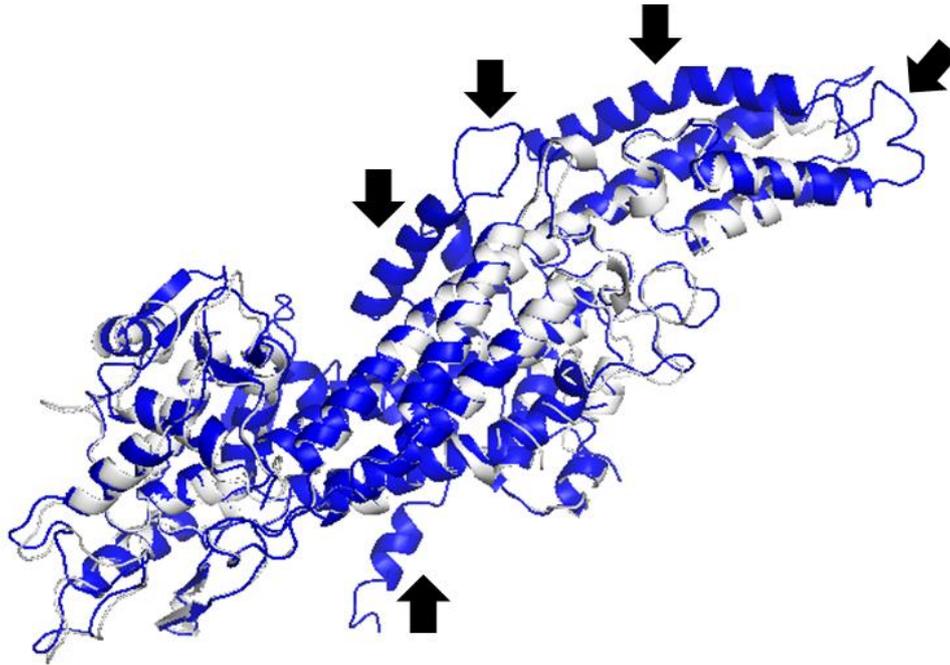


Figura 20. Alinhamento estrutural da mPAL_c1 (em branco) e 1w27_A (em azul). O alinhamento foi realizado no *software* PyMol. As setas em preto mostram zonas presentes na estrutura resolvida do PDB (1W27) e não mostradas ou mal alinhadas com nossa estrutura.

Tabela 13. Análise dos pockets da estrutura mPAL_c1.

<i>Pocket</i>	<i>Multiplicidade</i>	<i>Ligando(s)</i>	<i>Resíduos</i>
1	91	t-CA	H75, S135, L138, N187, N332, H344, F348, E430, N433, Q434
2	63	t-CA	T297, Q300
3	63	t-CA	K402, G403, I406

Wang *et al.* (2008) reportam que cisteínas expostas na superfície da proteína podem trazer à agregação. Verificou-se que as cisteínas C280, C466 e C516 ficam expostas mesmo após a formação do tetrâmero (Figura 19E), o que pode levar à agregação proteica, bem como instabilizar a catálise após temperaturas elevadas. O que é reforçado pela análise estrutural onde cerca de 8-13% da estrutura se manteria por ligações de hidrogênio, o que é muito pouco quando comparado ao resultado de outros autores, como Kovács *et al.* (2014) que observaram uma taxa de formação de pontes

salinas e ligações de hidrogênio próxima a 16%, o que promoveu a catálise sob temperaturas mais altas. Dessa forma, acredita-se que a mPAL_c1 seja de difícil expressão heteróloga e possa ser mais termolábil.

Encontrou-se também que as cisteínas C29, C296 e C412 da estrutura estão internalizadas, o que pode favorecer a formação do tetrâmero e sua estabilização, o que aumentaria sua estabilidade de catálise.

Análise filogenética

Para investigar a relação evolutiva entre mPAL_c1 e as sequências parentais de seu *core* catalítico, uma árvore filogenética foi construída pelo método de ML. Como mostrado na Figura 21, a árvore filogenética agrupou dois grandes ramos principais, incluindo espécies de plantas e fungos, segregando a sequência bacteriana como um *out group*. O resultado mostrou que mPAL_c1 é uma transição entre os grupos vegetal e fúngico, entretanto, com uma relação mais estreita com espécies vegetais. Uma discussão mais aprofundada sobre a relação das PAL parentais foi apresentada no capítulo 2. Um ponto interessante, é que os nossos resultados sugerem que mPAL_c1 compartilha um ancestral evolutivo comum com as PALs vegetais, o que quer dizer que o protocolo de *design in silico* proposto também pode ser utilizado para inferir configurações proteicas ancestrais, partindo dos pressupostos da homologia de aminoácidos e motivos conservados.

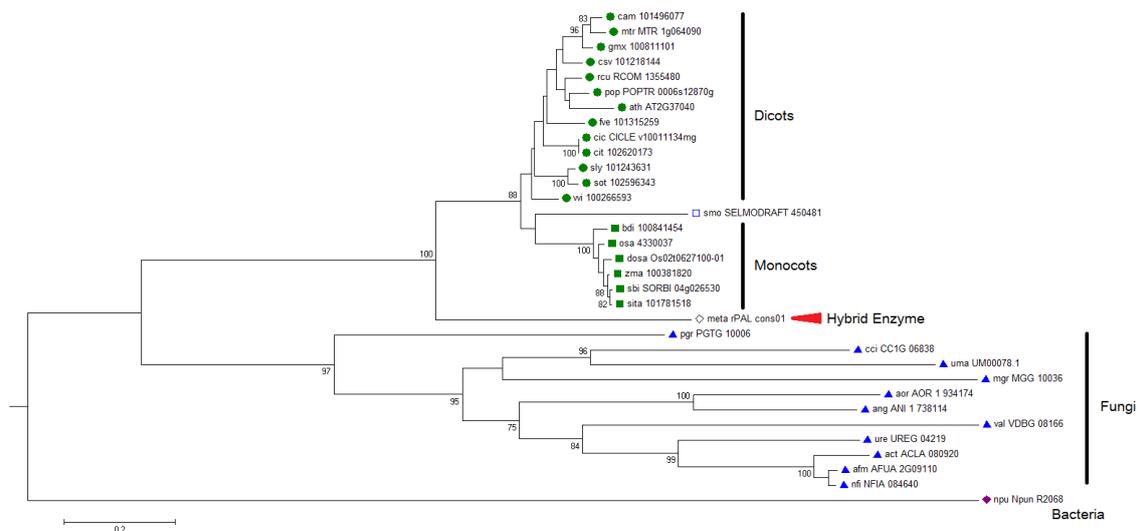


Figura 21. Análise filogenética com as PAL parentais do core catalítico da mPAL_c1. Para reconstrução filogenética utilizou-se o programa MEGA 6 e o método

ML. A árvore foi construída com o comprimento dos braços proporcional à distância evolutiva dada em substituições por sítio. A confiança estatística de cada nó foi dada a partir de um bootstrap com 1000 pseudorréplicas.

Expressão heteróloga

Foi possível observar a expressão de mPAL_c1 em hospedeiro *E. coli* Rosetta (DE3) a 37°C utilizando-se meio LB a partir de 1h de indução, entretanto, com 3h da indução obteve-se quantidades não muito superiores que nos tempos subsequentes, dessa forma considerou-se esse tempo como ótimo para produção (Figura 22). A produção de mPAL_c1 ocorreu na sua forma insolúvel em todos os cenários testados: mudança de temperaturas (37°C, 25°C, 15°C e 4°C); adição de compostos que auxiliam no enovelamento protéico, tais como altas concentrações de osmólitos (sacarose 5%); presença de altas pressões osmóticas (NaCl 0,5M); alteração do meio de cultivo (LB e *Terrific Broth* (SAMBROOK e RUSSEL, 2001)), quantidade de indutor, no caso, o IPTG (0,02 mM - 2 mM) e densidade da cultura na indução (OD₆₀₀ de 1,4 por 30 min ou OD₆₀₀ de 0,3-0,4 por 4 h). A insolubilidade completa da enzima foi confirmada no *Western Blot* realizado com anticorpo Anti-*His-Tag* (Figura 22). Como previsto na análise estrutural, provavelmente as cisteínas da superfície da proteína auxiliam na formação do agregado e o mantém estável, entretanto, não é viável eliminar o agregado durante sua produção, pois compostos que atuam reduzindo pontes dissulfeto, tais como 2-mercaptoetanol e DTT podem alterar a formação do cofator MIO, levando à inativação enzimática (WANG *et al.*, 2008; POPPE, 2001; CAMM e TOWERS, 1973).

Wang *et al.* (2005) enfrentaram um problema similar tentando expressar PAL de *Nostoc punctiforme* (um dos parentais da mPAL_c1) em *E. coli*. Para que houvesse produção de níveis significativos de NpPAL solúvel marcada com His-Tag foi necessária a co-expressão de chaperonas GroES/GroEL para facilitar o enovelamento correto da proteína no hospedeiro. Wang *et al.* (2008) trabalharam com uma PAL da alga *Anabaena variabilis* e a enzima possuía tendência a formar agregados, devido à cisteínas na superfície proteica. Por meio da mutação sítio dirigida das cisteínas externas, trocando-as por serinas, as proteínas mutantes resultaram em altas taxas de solubilidade. No entanto, apesar da facilidade dessas abordagens, sua realização não é nada prática. Isto, pois apesar da expressão envolvendo GroES/GroEL render grandes quantidades de proteína, o hospedeiro que optamos, *E. coli* Rosetta(DE3) já carrega um

plasmídeo codificante de códons-raros (pRARE) o que dificultaria ainda mais o processo, pois o plasmídeo *helper* contendo as chaperonas teria de ter uma marca de seleção alternativa à canamicina, ampicilina e cloranfenicol; além disto, um sistema celular mantido com 3 plasmídeos não parece estável e eficiente para experimentos de expressão heteróloga. Ainda há a impossibilidade de realização de mutações sítio dirigidas devido ao tempo e custo associados a tais técnicas. Dessa forma, optamos por obter a proteína por métodos de *refolding in vitro* haja visto que tais procedimentos podem render quantidades significantes de proteína e se revelam bastante úteis na prática.

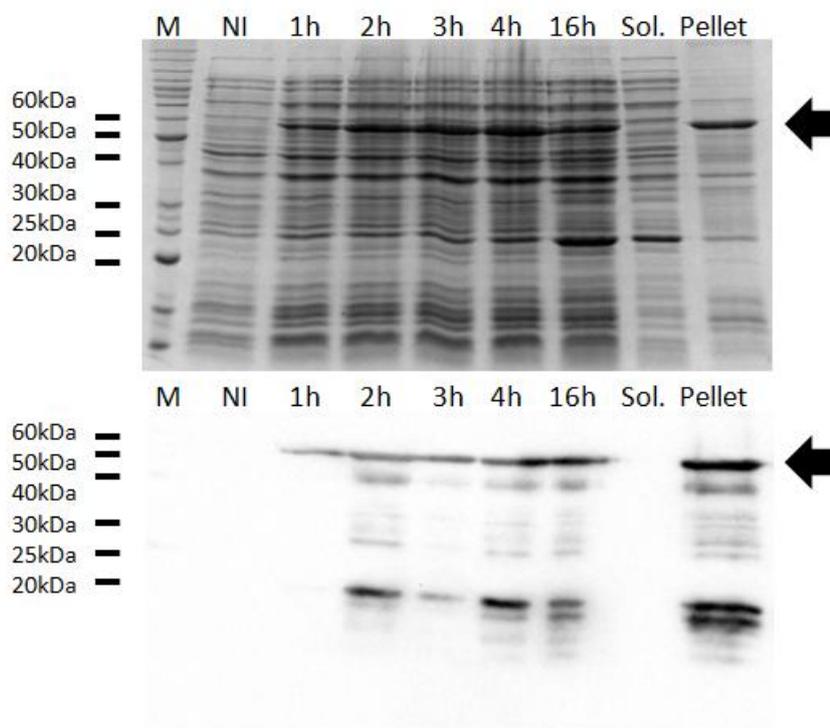


Figura 22. Expressão (acima) e Western Blot (abaixo) da mPAL_c1 recombinante. Um SDS-PAGE mostra as proteínas totais de *E. coli* Rosetta(DE3) recolhidas em 0h (NI), 1h, 2h, 3h, 4h e 16h pós-indução, sendo que as frações Sol. e *Pellet* representam a fração das proteínas totais solúvel e insolúvel após lise sob ultrassom e centrifugação, respectivamente. M representa o marcador *Bench Marker* (Invitrogen) e a seta preta mostra a altura esperada para mPAL_c1. O Western Blot do SDS-PAGE espelho mostrado foi realizado com marcação *anti-His-tag* e revelado com reagente ECL.

A uréia, frente aos outros agentes desnaturantes, apresenta diversas vantagens, tais como: baixo custo, facilidade de obtenção e manuseio, além de possuir um

mecanismo bastante conhecido. Ela, de modo geral, atua inicialmente desidratando a superfície da proteína pelo rompimento das pontes de hidrogênio estabelecidas com a água, facilitando a desnaturação, e num segundo momento, ligando-se preferencialmente à proteína através de interações dispersivas fortes, competindo com a camada de solvatação (FRANK e FRANKS, 1968; TOBI *et al.*, 2003; O'BRIEN *et al.*, 2007).

Após a ressuspensão do *pellet* (predominantemente mPAL_c1 insolúvel) e a adição de uréia até uma concentração final de 8M, deixamos a suspensão sob leve agitação a 4°C. O baixo nível de movimento mantém as partículas suspensas, evita a formação de agregados, auxiliando na solubilização das proteínas. Após a solubilização (Figura 24.A), os agregados restantes e contaminantes são separados por centrifugação e filtração, o que prepara a amostra para a IMAC em ambiente desnaturante em coluna de Ni²⁺-NTA. Nessa etapa há a eluição da proteína por meio da mudança de pH do meio através de lavagens com diferentes tampões com uréia 8M (Figura 23).

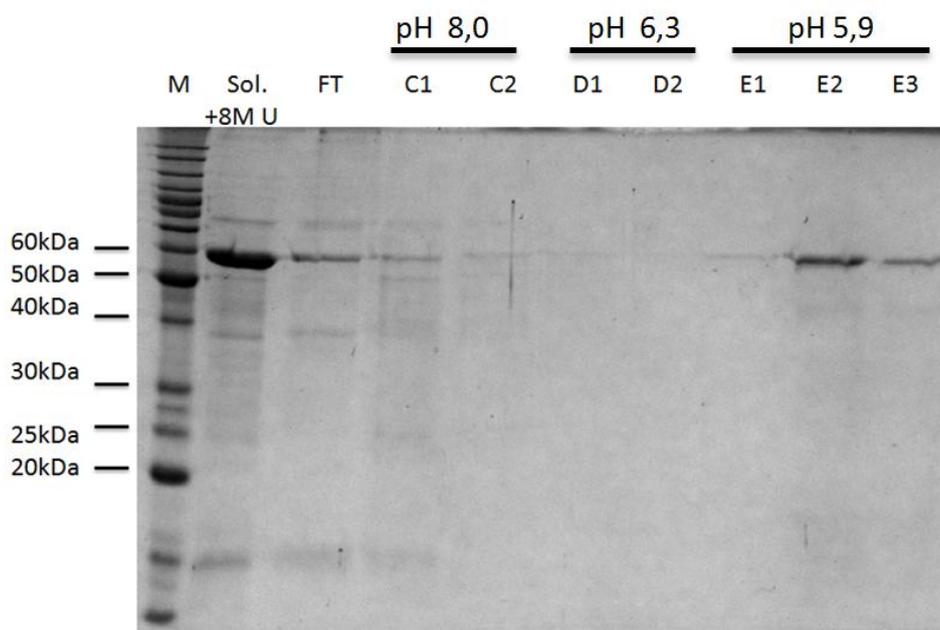


Figura 23. IMAC em ambiente desnaturante para purificação da mPAL_c1. A mPAL_c1 é liberada da resina à medida que há uma redução do pH e os grupamentos imidazólicos dos resíduos de histidina são protonados. "Sol.+8M U" refere-se à porção solúvel de proteína obtida após tratamento com uréia; FT refere-se à fração de proteínas não ligada à resina; C1-C2 refere-se à lavagens no pH 8,0; D1-D2 refere-se à lavagens no pH 6,3; E1-E3 refere-se à eluição no pH 5,9; M refere-se ao marcador *Bench Marker* (Invitrogen).

A partir do material obtido na purificação por IMAC em ambiente desnaturante procedeu-se com o *refolding in vitro* (Figura 24.B). Adotamos três técnicas: principais, ultrafiltração, diálise direta e diluição rápida. Adotamos ainda um parâmetro que consiste na relação de absorvâncias da amostra a 280 e 330 nm, conforme relatado por Poppe (2001) este é o comprimento de onda no qual o grupamento MIO ionizado absorve preferencialmente e mais fracamente quando neutro, e por se apresentar distante do comprimento de onda no qual as proteínas absorvem (280 nm), se revelou útil na avaliação da extensão de sua restituição.

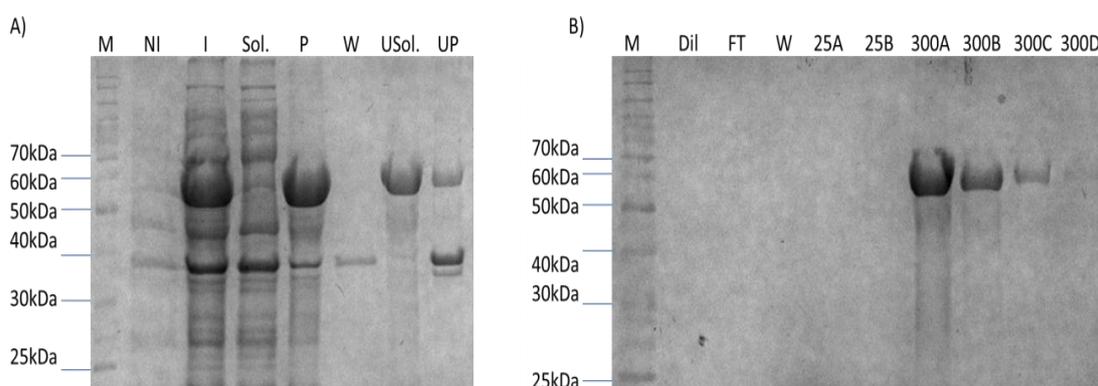


Figura 24. SDS-PAGE do processo de purificação da mPAL_c1. O processo de solubilização da mPAL_c1 com uréia (A) e a purificação da fração diluída da mPAL_c1 pré-purificada na segunda IMAC é mostrada em (B). A pureza da proteína final é mostrado nas frações de eluição final 300B a 300D. Legenda: M - marcador *Bechn Marker* (Invitrogen), NI - amostra não induzida, I - amostra induzida por 3h, Sol. - fração solúvel, P - fração insolúvel, W - lavagem do *pellet*, USol. - fração solúvel após tratamento com uréia, UP - fração insolúvel após tratamento com uréia, Dil. - amostra diluída após purificação em IMAC desnaturante, FT - fração não ligada à coluna, W - lavagem da coluna, 25 A e B - eluição contendo imidazol 25 mM, 300 A a D - eluição contendo imidazol 300 mM.

Na Tabela 14 é possível observar que o método de ultrafiltração possui baixo rendimento e a enzima obtida por este método não apresentava nenhuma atividade detectável com L-Phe. O padrão de pureza das amostras renaturadas *in vitro*, pelos métodos testados, evidenciado por SDS-PAGE 12% mostra que o método de ultrafiltração também foi ineficiente por concentrar impurezas da amostra original

(dados não mostrados). Nesse cenário, estudos preliminares confirmaram que ambos os métodos de diluição rápida e diálise direta mantinham a atividade da mPAL_c1, entretanto, a atividade da enzima resultante da diluição rápida era 2 vezes maior e o método rendia aproximadamente a mesma quantidade de enzima. Dessa forma, optou-se pela produção da enzima por diluição rápida (Figura 24.B). Este trabalho, é o primeiro, do nosso conhecimento, que utiliza a técnica de renaturação *in vitro* com a enzima PAL.

Tabela 14. Resumo das técnicas de *refolding* utilizadas na mPAL_c1. A razão de absorvâncias 280/330 mostra proporção da proteína total que possui o cofator MIO restituído.

Processo	Abs _{280/330}	mg/L de cultura	Atividade específica (U.mg ⁻¹)
Diálise direta	1,33	5,0	0,86
Diluição Rápida	1,17	4,7	1,5
Ultrafiltração	1,21	1,5	0,0

Qualidade estrutural da mPAL_c1

Após a produção e purificação da mPAL_c1, testes de qualidade estrutural foram realizados a fim de verificar a conformação final da enzima devido aos processos de renaturação *in vitro*. Inicialmente, testamos a eficiência dos métodos de diluição rápida e de diálise direta na reconstituição do cofator cataliticamente essencial MIO (Tabela 14, Figura 25) por meio da varredura da absorção em comprimentos de onda característicos do UV-Próximo (RÖTHER *et al.*, 2000). Os resultados mostraram claramente que o método de *refolding* por diluição rápida foi capaz de restituir o cofator MIO de modo mais eficiente, pois as amostras provenientes desta técnica apresentam picos característicos na faixa de 310 nm e 330 nm (POPPE, 2001), que são superiores àqueles apresentados pelas amostras renaturadas via diálise direta. Outros autores relatam que o pico máximo para absorção de MIO é de 313 nm (CHRISTENSON *et al.*, 2003; RÖTHER *et al.*, 2000), que também foi observado nas amostras obtidas por diluição rápida, diferentemente das amostras por diálise direta.

A absorvância a 338 nm das amostras obtidas por diálise direta foi 3 vezes maior do que as obtidas por diluição rápida, mostra que provavelmente a atividade reduzida e a regeneração do cofator MIO foram afetados por elementos que entraram em contato com a amostra e atuaram nesse sítio. Nesse caso, agentes redutores de sulfidrilas ou propriamente cisteínas podem ter interagido no sítio do cofator MIO gerando este sinal, o que significa enovelamento incorreto.

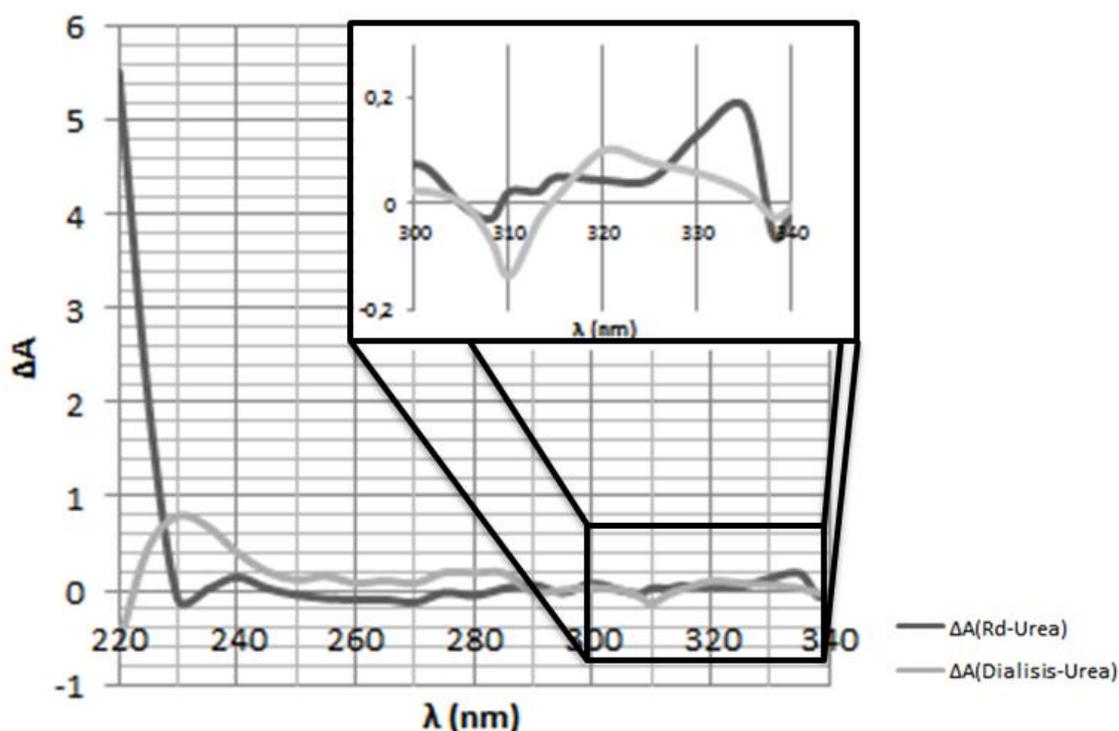


Figura 25. Análise da reconstituição do cofator MIO (4-metilidenoimidazol-5-ona) depois do processo de refolding. Todas as curvas foram feitas usando o respectivo tampão como branco e foram dadas no gráfico final como curvas diferenciais. A região de 300 a 340 nm foi ampliada no canto direito superior. Legenda: Rd - urea, curva da amostra renaturada pelo método de diluição rápida menos a curva obtida com a amostra desnaturada com 8 M de uréia; Diálise - urea, curva da amostra renaturada pelo método da diálise rápida menos a curva obtida com a amostra desnaturada com 8 M de uréia.

Os métodos de *refolding* dependem em sua maioria da manutenção de um estágio intermediário de equilíbrio do *fold*ing que existe nas concentrações médias de desnaturante. Tais estruturas intermediárias são instáveis e menos solúveis e, como

consequência podem gerar agregados e *misfolds* proteicos (TSUMOTO *et al.*, 2003). Dessa forma, manter a solubilidade e flexibilidade do intermediário pode garantir uma estrutura final nativa estável. A forma como o reenovelamento das proteínas ocorre depende intrinsecamente da forma como o agente desnaturante tem sua concentração reduzida na solução. Sob altas concentrações de agentes desnaturantes, as proteínas perdem sua organização estrutural, e se tornam estruturas lineares, bem solvatadas e flexíveis. Em tampões aquosos, por sua vez, elas atingem estágios de enovelamento tais que se tornam glóbulos ou fibras, rígidos e compactos (RICHARDS, 1974). Idealmente, a transferência de moléculas de proteína de alta concentração de desnaturante para tampão aquoso deve ser cuidadosamente realizada para que haja a transferência de moléculas do agente desnaturante na superfície da proteína para o solvente, forçando assim o colapso da estrutura desordenada, numa estrutura compacta (SINGH e PANDA, 2005).

Apesar dos três métodos testados, apenas o método da diluição rápida não apresentou agregados visíveis. Acreditamos que o método da diálise direta possa ter sido eficiente até passos anteriores à completa eliminação da uréia (de 0,5M de uréia até 0M), e essa transição brusca de solvente possa ter gerado uma instabilidade na forma intermediária do enovelamento. Entretanto, esses resquícios de desnaturante poderiam afetar a atividade enzimática final, portanto, técnicas que eliminem traços desses agentes desnaturantes da amostra são preferidas, tais como a diluição rápida (YANG *et al.*, 2011).

Os agregados proteicos podem ocorrer numa infinidade de tamanhos, e uma vez agregados ou mal renaturados, na ausência de moléculas desnaturantes, não têm flexibilidade para desagregar e se dobrar na estrutura nativa (SINGH e PANDA, 2005; TSUMOTO *et al.*, 2003). Dessa forma, optamos por confirmar o estado oligomérico e de agregação proteica nas amostras obtidas por diluição rápida por meio de exclusão molecular (Figura 26). Este experimento demonstrou que a proteína se apresenta em alto grau de pureza, e primariamente em estado agregado, haja visto, o pico protéico principal ser eluído juntamente com o volume "morto" (V_0) da coluna, mas a amostra ainda possui formas aparentemente diméricas e monoméricas da mPAL_c1 associadas em uma menor extensão. Essas formas aparentemente monoméricas e diméricas, não apresentam atividade enzimática associada e podem representar parcelas de proteína reenovelada corretamente.

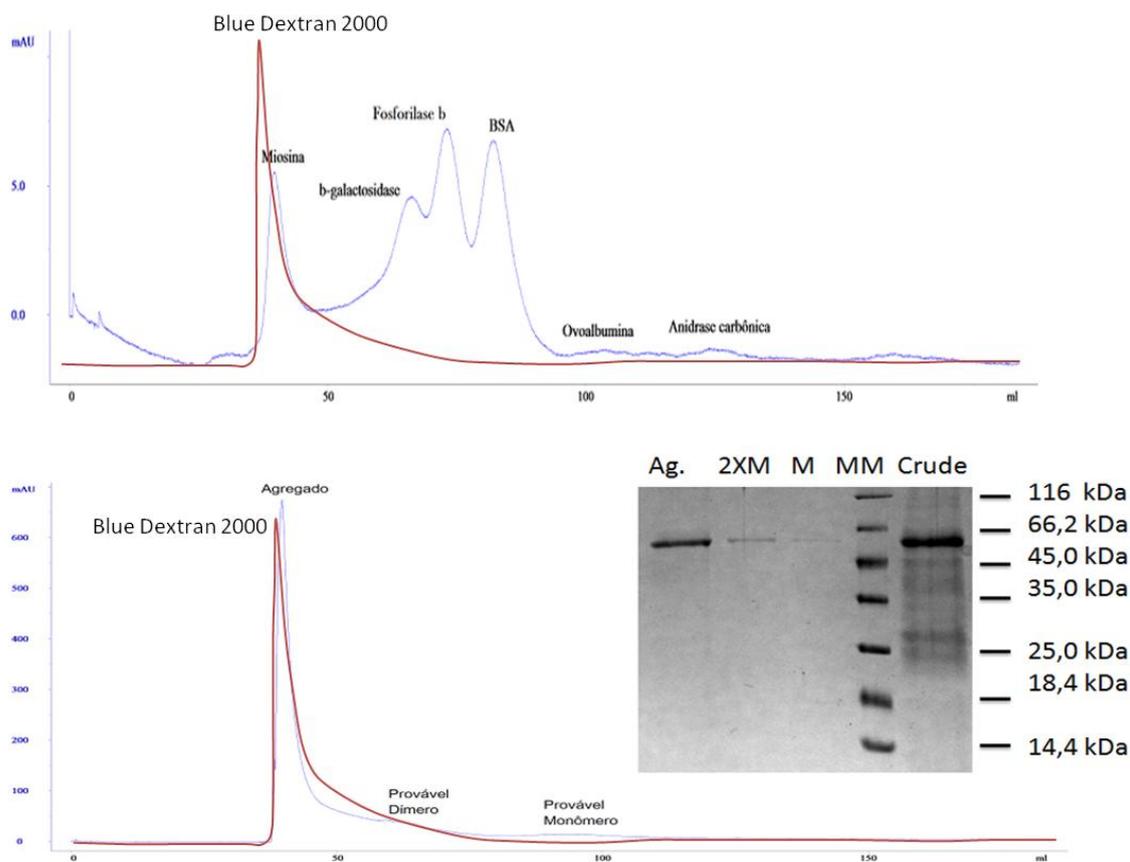


Figura 26. Cromatografia de exclusão molecular da mPAL_c1 renaturada. A exclusão foi feita numa coluna HiPrep Sephacryl S-300 HR 16/60. Todos os métodos cromatográficos foram realizados usando o tampão de diálise (25 mM de NaH_2PO_4 , 50 mM de NaCl, pH 8,0), incluindo ensaios de calibração. A curva em vermelho mostra a corrida com Blue Dextran 2000, utilizado para determinação do volume "morto" (V_0). O cromatograma superior mostra a curva de calibração, enquanto que o cromatograma inferior mostra a corrida com a mPAL_c1 renaturada. O SDS-PAGE foi realizado utilizando-se amostras de cada pico obtido na separação da mPAL_c1. Legenda: Ag. - agregado de mPAL_c1, 2xM - homodímero aparente de mPAL_c1, M - monômero aparente de mPAL_c1, MM - *PierceTM Unstained Protein MW marker* (Thermo), Crude - amostra inicial desnaturada com uréia e pré-purificada por IMAC em condições desnaturantes.

Com base nos dados levantados na cromatografia de exclusão molecular pudemos calcular a massa molecular e parâmetros hidrodinâmicos relacionados à forma da proteína (Tabela 15). Diferentemente do senso comum, a exclusão molecular não

separa moléculas por massa e sim por forma e tamanho, o que se reflete no raio hidrodinâmico (Rh). Assim, estimativas da massa molecular realizadas com base na exclusão molecular podem sofrer erros, isto porque geralmente as proteínas utilizadas no processo de calibração da coluna são globulares, e caso a proteína a ser determinada se afaste da forma globular, pode haver um desvio no volume de eluição que acarreta um erro de até 2 vezes do tamanho esperado (ERICKSON, 2009). Assim sendo, utilizamos o parâmetro ideal para comparações que é o Rh (SIEGEL e MONTE, 1966), que está diretamente ligado à forma da proteína e garante que mesmo um elipsóide oblongato como a mPAL_c1 possa ser comparado com os padrões globulares.

Tabela 15. Principais características hidrodinâmicas da mPAL_c1 calculadas pelos dados da cromatografia de exclusão molecular. Os parâmetros hidrodinâmicos obtidos foram resumidos da seguinte forma: MW refere-se a massa molecular em kDa; Rh refere-se ao raio de Stokes ou raio hidrodinâmico em nm; R_0 refere-se ao raio de uma esfera anidra de mesma massa em nm (calculado pelo Sednterp); Rh/ R_0 refere-se a razão friccional; (%) refere-se à abundância relativa de espécies na amostra inicial.

mPAL_c1	MW (kDa)		(nm)		Rh/ R_0	(%)
	Calculado	Medido	Rh	R_0		
Monômero Aparente	58	63	4,5	2,6	1,7	0,1
Dímero Aparente	116	109	6,6	3,2	2,1	0,1
Agregado	233	214	12	4,1	2,9	98

Os dados calculados (Tabela 15) são importantes para determinação de 2 parâmetros: forma e massa da proteína. A massa da proteína pode ser estimada com precisão devido ao conhecimento prévio de sua sequência e da incapacidade de bactérias realizarem modificações pós-traducionais. Ainda sabendo que a proteína é uma molécula assimétrica (Figura 19) temos que o erro na determinação direta da massa poderia ser de até 50% (ERICKSON, 2009), entretanto, os erros obtidos não chegam a 9%. Esses desvios são inferiores aos maiores desvios de até 33% reportados por outros autores (HESTERBERG *et al.*, 1981). Vale ressaltar que uma das proteínas de calibração, a miosina (205 kDa) eluiu em volume muito próximo ao do Blue Dextran 2000, o que fez com que ela fosse excluída da curva padrão, dessa forma só podemos

supor que a porção que eluiu como "agregado" tenha massa superior a 200 kDa e que a aproximação da massa dos picos similares ao dímero e monômero dos valores teóricos apesar de sua forma irregular é uma consequência disso. O Blue Dextran 2000 tem se mostrado não recomendável para determinação de V_0 , pois é heterogêneo e dá resultados variáveis, atualmente tem se utilizado moléculas grandes como a tiroglobulina para tal finalidade (Bio-Works Application nº 301), no entanto, muitos laboratórios ainda utilizam sistemas dependentes dele.

Para determinação da razão friccional (R_h/R_0), o raio hidrodinâmico de uma esfera totalmente lisa e anidra de massa equivalente à da proteína (R_0) foi calculado pelo *software* Sednterp (<http://sednterp.unh.edu/>). A razão friccional de uma proteína mede quão assimétrica ela é, dessa forma se a massa ou a assimetria da proteína são aumentados, ele tende a aumentar, por outro lado, se considerarmos que a proteína seja uma esfera lisa sem água ligada teremos então seu valor mínimo. A razão friccional é análoga à relação entre os coeficientes de sedimentação máximo calculado e medido ($S_{Máx}/S_{20,w}$), e ambos resultam em valores próximos que podem ser utilizados para predição da forma da molécula (SCHÜRMAN *et al.*, 2001; ERICKSON, 2009). Assim as formas de mPAL_c1 possivelmente identificadas foram: o monômero aparente como uma proteína moderadamente alongada (faixa de 1,5 a 1,9), enquanto o dímero aparente e o agregado como proteínas altamente alongadas (faixa de 2,0 a 3,0), similarmente à tropomiosina ou o fibrinogênio, por exemplo. Essas informações são consistentes com o que observamos anteriormente no modelo tridimensional da proteína que foi proposto (Figura 19).

As formas monomérica e dimérica da enzima não apresentaram atividade enzimática, assim passamos a trabalhar com o agregado que apresentou sinais de atividade muito baixos. Testes envolvendo a estrutura secundária e terciária do agregado foram então realizados. Os testes da emissão de fluorescência do triptofano são largamente utilizados para verificação de estruturas enoveladas (TCHERKASSAKAYA *et al.*, 2000). Eles se baseiam na alteração do comprimento de onda da emissão de resíduos de triptofano devido à polaridade do ambiente em que eles estão (BEECHEM e BRAND, 1985; DEMCHENKO, 1986; EFTINK, 1991; KONEV, 1967). Trp $\lambda_{Máx}$ geralmente varia de 308 nm (azurina) a 355 nm (por exemplo, Glucagon) e é correlacionado com a exposição do cromóforo ao solvente (VIVIAN e CALLIS, 2001).

Foi realizado o teste de emissão utilizando para tal uma amostra completamente desnaturada em uréia 8M e outra renaturada pelo processo anterior (Figura 27).

Verificou-se um desvio das curvas do estado desnaturado e do estado renaturado de 5 nm no $\lambda_{M\acute{a}x}$ (345 \rightarrow 340 nm), representando um deslocamento hipsocrômico. Segundo o modelo de estado discreto de Burstein (BURSTEIN *et al.*, 1973; RESHETNYAK e BURSTEIN, 1997), determinamos que houve uma transição de ambientes por meio da alteração do $\lambda_{M\acute{a}x}$: de um estado II/III em que a cadeia lateral do resíduo está exposta na superfície da proteína em contato com águas móveis e imobilizadas ($\lambda_{M\acute{a}x} \sim 342\text{-}350$ nm), para o estado I, onde há um exciplex (2:1) no interior da proteína ($\lambda_{M\acute{a}x} \sim 331\text{-}340$ nm).

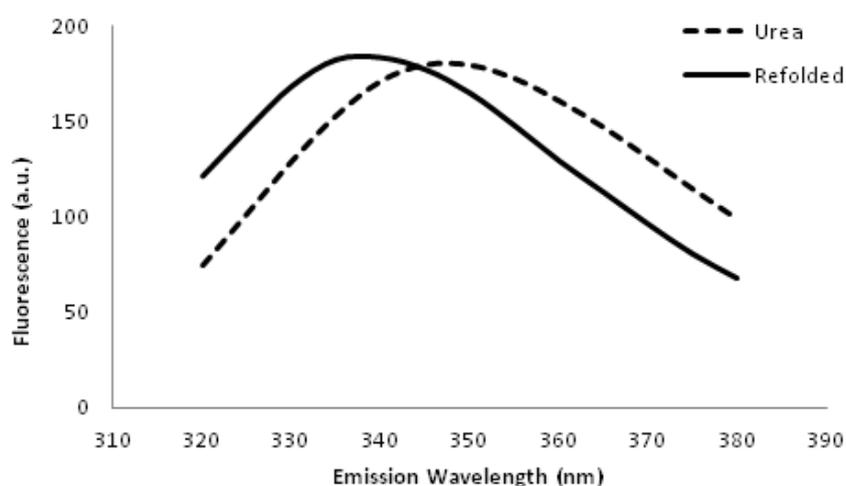


Figura 27. Ensaio da emissão de fluorescência do Trp. A emissão do triptofano foi escaneada por vários comprimentos de onda, a partir da excitação a 295 nm. O ensaio foi realizado a 20°C e utilizou-se 5,8 μM de mPAL_c1 desnaturada e renaturada para construção das curvas "urea" e "refolded", respectivamente. O desvio hipsocrômico entre os $\lambda_{M\acute{a}x}$ mostra que os triptofanos foram parcialmente escondidos, indicando variação no *folding* da mPAL_c1.

Considerando que mPAL_c1 possui 2 resíduos de Trp e um deles está exposto na superfície e outro está incorporado no interior do possível tetrâmero (Figura 19.D). A renaturação parcial da enzima foi eficientemente evidenciada pelo ensaio de emissão de

fluorescência. Portanto, podemos supor que a enzima enquanto estado agregado se enovelou internalizando parte dos triptofanos num ambiente hidrofóbico, a outra parte acaba desviando a emissão para comprimentos de onda maiores devido à interações com a água de hidratação (*red shift*) (VIVIAN e CALLIS, 2001).

A verificação de estrutura secundária foi realizada por meio de dicroísmo circular. Esta técnica se baseia na absorção desigual dos componentes L e R da luz por centros quirais, e o ângulo resultante desta absorção diferencial é medido no espectropolarímetro (KELLY *et al.*, 2005). O resultado da análise realizada é mostrado na Figura 28A.

A relação entre as elipsidades molares residuais entre os comprimentos de onda de 222 nm e 200 nm revela padrões de clusterização bastante específicos (WIEDEMAN *et al.*, 2013), sabendo-se que estruturas desnaturadas tendem a possuir um mínimo a 200 nm e aproximadamente 0 a 222 nm. Após um treinamento de redes neurais, obteve-se predições dos estados de enovelamento da proteína (Figura 28B) que mostrou o agregado mPAL_c1 como uma proteína de enovelamento intermediário entre a forma nativa e de glóbulo pré-fundido. Kelly *et al.* (2005) recomendam que durante a deconvolução de curvas do CD sejam testados mais de um sistema, por isso, durante a deconvolução da curva, testou-se os programas: CDNN (Böhm, 1997), resultados completos disponíveis no ANEXO VIII, e o servidor CAPITO (<http://capito.nmr.leibniz-fli.de/>) com as configurações em *default*. Os resultados obtidos são mostrados na Tabela 16.

A forma real e magnitude do CD no espectro do UV próximo de uma proteína irá depender do número de cada tipo de aminoácido aromático presente (KRITTANAI *et al.*, 1997), a sua mobilidade, a natureza do seu ambiente e a sua disposição espacial na proteína. Centros quirais vicinais próximos (distância inferior a 1 nm de intervalo) podem ser capazes de parrear seus sinais de excitação, apesar de os sinais serem fracos demais para que isso seja significativo (KELLY *et al.*, 2005). Assim, o sinal do agregado no CD pode ter sido desviado por diversos fatores além da presença do grupamento cromóforo MIO.

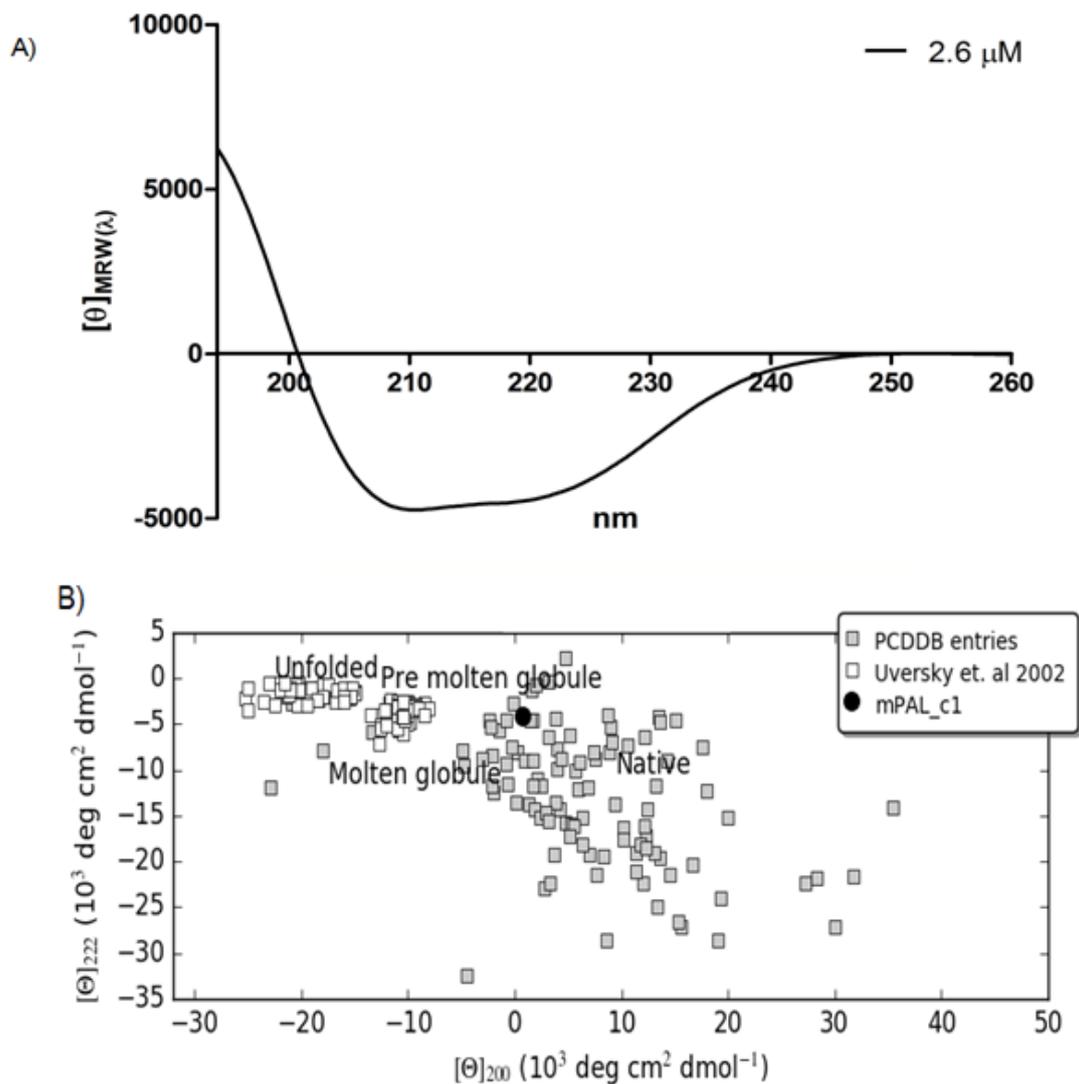


Figura 28. Dicroísmo circular do agregado da mPAL_c1. São mostrados a curva de absorptividade residual média do agregado mPAL_c1 (A) e a relação entre a absorptividade residual média nos comprimentos de 200 e 222 nm (B) juntamente com um diagrama contendo os valores de diversas proteínas já descritas, enoveladas e desnaturadas, obtido no servidor CAPITO. Ambos os gráficos revelam que o agregado mPAL_c1 está no estado de transição entre a forma nativa e de glóbulo pré-fundido.

Tabela 16. Estrutura secundária do agregado mPAL_c1 por dicroísmo circular sob deconvolução em diferentes softwares. Os valores esperados foram obtidos do modelo tridimensional da Figura 19.

Estrutura	CDNN	CAPITO	Esperado
Hélices	11%	13%	54%
Folhas β	36%	42%	3%
Beta-Turns	19%	51%	32%
Rndm. Coils	35%		11%
Total	101%	105%	101%

O CD tem sido utilizado para medir a taxa de aquisição de estrutura secundária e terciária. No entanto, o enovelamento de proteínas maiores (como no caso da mPAL_c1), muitas vezes envolve uma via multi-etapas, com formas intermediárias nas fases iniciais em que a estrutura secundária semelhante à nativa é adquirida e antes da formação de contatos corretos para formação da estrutura terciária (TSUMOTO *et al.*, 2003). Estes intermediários iniciais são ditos "glóbulos fundidos" (RADFORD, 2000; JEMTH *et al.*, 2004; DOBSON, 2004), dessa forma nossos resultados sugerem que o agregado mPAL_c1 seja uma estrutura intermediária (Figura 28B). Protocolos de *refolding* geralmente utilizam agentes redutores como o β -mercaptoetanol para evitar agregação por ligações sulfidrílicas, no entanto, seu uso é limitado nesse caso, pois a ação deste agente e de agentes similares pode comprometer a atividade enzimática de modo irreversível, já que o cofator MIO tende a ser inativado por ele.

A mPAL_c1 teoricamente possuiria um núcleo central de hélices paralelas de diferentes comprimentos, e apenas uma seção de folhas β maior que 3 resíduos; essa conformação geraria um funil, que desvia as cargas e gera um momento dipolo tornando o sítio ativo apto à reação. Por isso o alto conteúdo de hélices esperado na estrutura final (54%), no entanto, o que observamos foi apenas 1/5 do conteúdo de hélices esperado, o que pode afetar sua taxa de catálise de modo significativo.

Esse enovelamento incorreto pode ter sido causado pelo processo de *refolding*, uma vez que a estrutura secundária de proteínas é principalmente determinada pela sequência de aminoácidos e estabilizada por ligações de hidrogênio. Sob condições de desnaturação, as proteínas com várias organizações de estrutura secundária semelhante

podem agregar-se em fibrilas amilóides ricas em folhas β (DING *et al.*, 2003). A interação helicoidal é abolida no estado desnaturado, enquanto a interação hidrofóbica não nativa entre as motivos internos da proteína podem ocorrer intra-molecularmente ou inter-molecularmente, facilitando ainda mais a formação de agregados (TSUMOTO *et al.*, 2003), como o da proteína mPAL_c1.

Atividade enzimática do agregado mPAL_c1

A atividade enzimática do agregado mPAL_c1 foi monitorada sob diferentes condições (Figura 29). Apesar de reportado na literatura que as PALs e HALs sofrem efeito de cátions divalentes diferentes (Tabela 17) (ASANO *et al.*, 2004), o efeito do cátion Mn^{2+} se revelou pouco influente sobre a atividade da enzima mPAL_c1, independentemente do substrato, por isso concentrações mínimas (2 mM) foram utilizadas nas reações (dados não mostrados).

Tabela 17. Efeito de diferentes cátions divalentes sobre PALs isoladas de diferentes fontes (Extraído e adaptado de BRENDA-Enzyme, 2016).

Metal	[mM]	Efeito	Organismo
Ca(II)	1-5	Estabilização	<i>R. glutinis</i>
Cd(II)	2	Ativação	<i>G. max</i>
Cs(II)	1	Ativação	<i>C. Deserticola</i>
Li(I)	1	Estabilização	<i>R. glutinis</i>
Mg(II)	8	Ativação	<i>C. pepo</i>
Mn(II)	8	Ativação	<i>C. pepo</i>
Mn(II)	2	Estabilização	<i>R. glutinis</i>
Ni(II)	0,4	Indução	<i>J. curcas</i>
Pb(II)	0,1	Estabilização	<i>G. max, L. luteus</i>
Zn(II)	1	Ativação	<i>R. glutinis</i>

A atividade da mPAL_c1, com L-Phe como substrato, apresentou um máximo num pH de 7,5 com 2mM de $MnCl_2$ (Figura 29.A). Já utilizando L-His como substrato, a produção de ácido urocânico pela mPAL_c1 teve seu máximo num pH alcalino (pH 12,0), como mostrado na Figura 29.B. A atividade da mPAL_c1 com L-Phe é estável

nos pHs entre 6,5-9,0 durante 1 h à temperatura ambiente. Em valores de pH extremo a atividade cai, provavelmente devido à desnaturação da proteína. O agregado mPAL_c1 é mais eficiente numa faixa de pH menor do que as das PALs até então descritas, que tem atividade em pHs alcalinos, geralmente com pH ótimo entre 8,5-9,5 (KOVÁCS *et al.*, 2014); dessa forma o agregado mPAL_c1 é interessante para reações em pHs mais neutros.

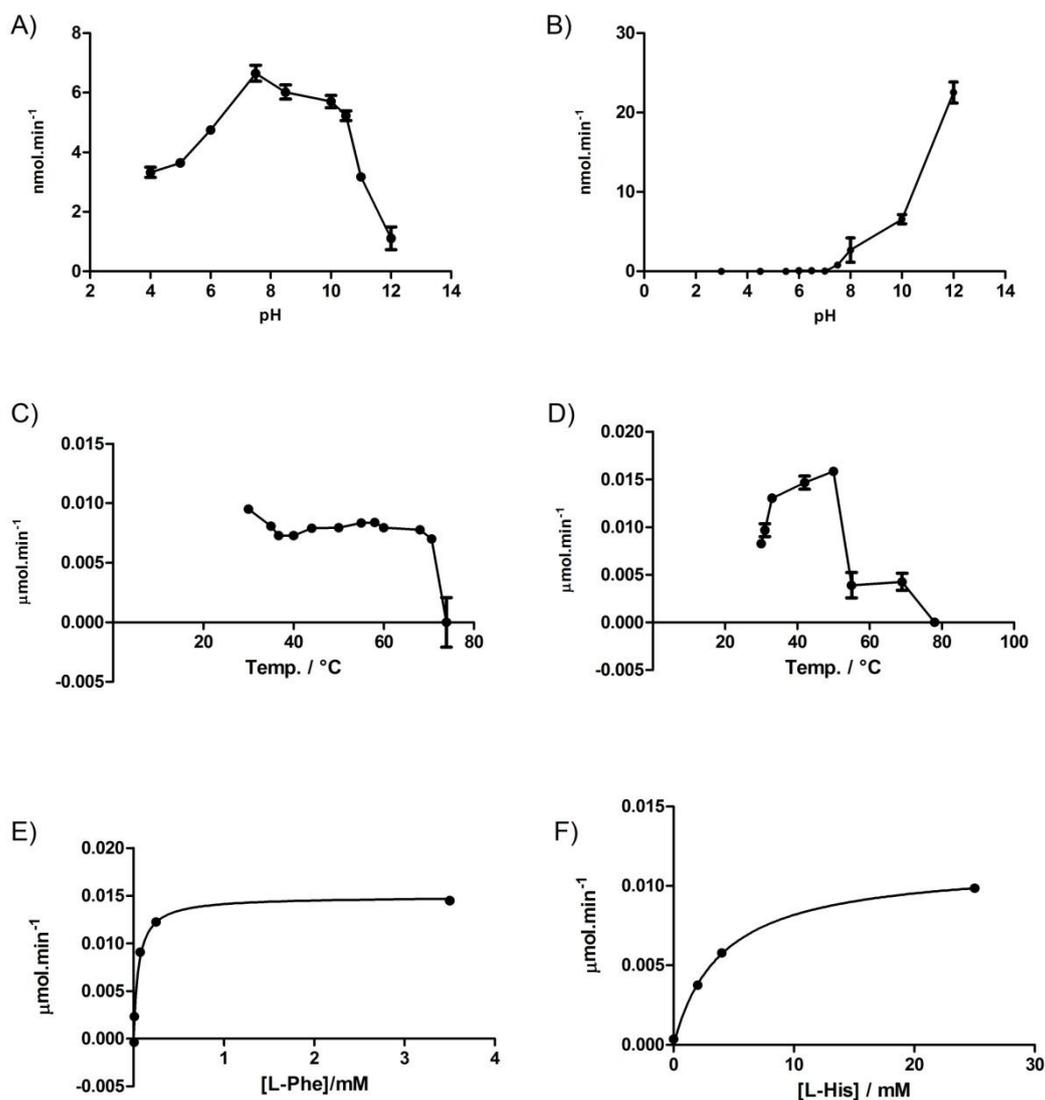


Figura 29. Ensaio de atividade da mPAL_c1. As atividades em diferentes pHs foram testadas com L-Phe (A) e L-His (B) como substrato. A temperatura foi avaliada tendo como substratos L-Phe (C) e L-His (D). A cinética foi feita utilizando-se diferentes concentrações de substrato no pH e temperatura ótimos previamente determinados (E-F)

e seguem a cinética Michaeliana. Os gráficos foram produzidos no *software* GraphPad Prism v.5.

O agregado mPAL_c1 se mostrou relativamente termorresistente como PAL, com uma perda de atividade de 50% a partir de 35°C, mantendo esta taxa de reação a temperaturas de até 65°C (Figura 29.C). Possui ainda um pico de atividade de HAL a 45°C (Figura 29.D). Determinou-se a $T_{opt.}$ de 30°C e 45°C para atividades de PAL e HAL, respectivamente. A perda de atividade do agregado mPAL_c1 sob temperaturas acima de 50°C parece ser acelerada por processos de agregação proteica, visualmente detectáveis. Um estudo que compreendeu mais de vinte enzimas, incluindo a PAL de *Rhodotorula glutinis* (PAL_Rho) indicaram que a sua estabilidade térmica foi correlacionada (embora fracamente) com a temperatura de crescimento do organismo fonte (DANIEL *et al.*, 2008). Kovács *et al.* (2014) descobriu que as interações iônicas mantêm a estrutura das PAL ativas em maiores temperaturas, e são responsáveis pela termorresistência de enzimas como a codificada pelo gene EncP de *Streptomyces maritimus* (CHESTERS *et al.*, 2012). Acreditamos que o conteúdo de aminoácidos polares da mPAL_c1, capazes de formar pontes salinas (21,4%) e ligações de hidrogênio (29,2%), sejam responsáveis pela sua termolabilidade, uma vez que as enzimas estudadas pelos autores anteriormente citados mostraram-se muito mais estáveis com o aumento das pontes salinas na estrutura final, reforçando a discussão do balanço de aminoácidos polares carregados iniciada no Capítulo 2.

A cinética do agregado mPAL_c1 com L-Phe possui K_M de 55 μ M e K_{cat} de 0,001 s^{-1} (Figura 29.E). A relação K_{cat}/K_M foi de 0,01633 $mM^{-1}s^{-1}$. Esses resultados sugerem que a atividade de PAL do agregado mPAL_c1 é aproximadamente 4,8 vezes menos eficiente que a enzima EncP de *S. maritimus* (K_{cat} 0,0048 s^{-1} , K_{cat}/K_M 0,21 $mM^{-1}.s^{-1}$ a 30°C, pH 8,8) (McGRATH *et al.*, 2006). Isso faz sentido, pois num agregado uma pequena porção de enzimas deve ser ativa, em detrimento da grande maioria, para cálculos de K_{cat} tais enzimas inativas são consideradas como ativas, e por isso, o valor é subestimado. Os valores de K_{cat} do agregado mPAL_c1 foram observados como o menor valor do banco de dados BRENDA-Enzyme (2016) (Figura 30.B), mostrando que sua atividade é comparada à atividade enzimática basal (Figura 30 A-B). Enzimas ancestrais são comumente associadas com baixas atividades e *bursts* catalíticos (NEWTON *et al.*, 2015), sugerindo que o agregado mPAL_c1 tenha uma natureza

ancestral que pode ser modificada a partir de experimentos de evolução *in vitro*, inclusive para melhor solubilidade e enovelamento. O agregado mPAL_c1 tendo como substrato L-His possuiu K_M de 2,609 mM com K_{cat} de $0,001\text{ s}^{-1}$ e K_{cat}/K_M de $0,230\text{ }\mu\text{M}^{-1}\text{ s}^{-1}$ (Figura 29.F). Todos estes resultados sugerem um resquício da estrutura receptora sob a forma de atividade de histidina amônia-liase.

As atividades reduzidas do agregado mPAL_c1 se devem evidentemente a 3 fatores:

- a agregação proteica, ocorrida por diversos fatores, e que pode ocluir os sítios ativos da enzima ou mantê-la numa conformação não favorável;
- nas moléculas de enovelamento correto pode ocorrer que o *loop* contendo a Tyr conservada, oclua o sítio ativo da mPAL_c1 e desordene o substrato, reduzindo o fluxo das moléculas de substrato para o *pocket*;
- a estrutura da enzima não seja favorável a altas taxas catalíticas.

Dado que o comprimento e estabilidade do *loop* são variáveis em diversas PALs, parece provável que a eficiência da abertura e o fechamento do *pocket* ativo, bem como a estabilidade da conformação fechada em relação a aberta também irá influenciar K_M e K_{cat} (WANG *et al.*, 2008).

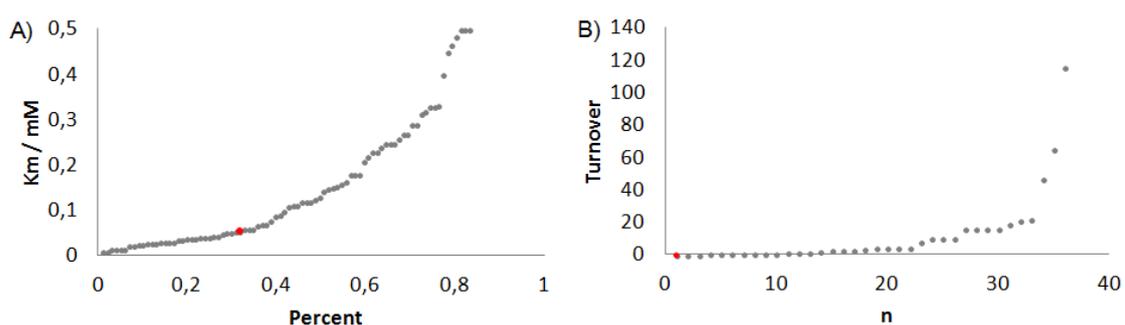


Figura 30. Constantes cinéticas comparadas do agregado mPAL_c1 com as PAL do banco de dados BRENDA-ENZYME. Os valores de K_M e K_{cat} disponíveis para cada entrada de PAL no BRENDA-Enzyme (2016) foram utilizados na comparação das constantes cinéticas da mPAL_c1. Os valores de K_M foram organizados e distribuídos em percentis (A), enquanto que os valores de K_{cat} foram distribuídos em posições de

ranking (B), os valores obtidos do agregado mPAL_c1 são mostrados como círculos vermelhos.

Há relatos na literatura de ativação da PAL pelo substrato (CAMM e TOWERS, 1973). Assim, a concentração de L-Phe foi aumentada de 2,0 a 5,0 mM, e a atividade foi avaliada (Figura 31). Observou-se um efeito de ativação do agregado mPAL_c1 de quase 5 vezes, neste intervalo, entretanto a PAL_Rho comercial apresentou uma ativação de quase 40 vezes. Quando comparado com a enzima comercial, PAL_Rho de *R. glutinis*, com 2 mM de L-Phe a atividade da mPAL_c1 foi quase três vezes inferior (cerca de 30% da atividade). Isso mostra que a enzima comercial é muito mais eficiente que o agregado mPAL_c1.

Uma das principais aplicações da PAL na indústria é sua utilização na produção de L-Phe a partir de t-CA pela sua reação inversa (YAMADA *et al*, 1981; HAMILTON *et al*, 1985; EVANS *et al*, 1987a, 1987b), entretanto, para que isso ocorra são necessárias altas concentrações de amônio, foi relatado que a partir de 4M há a inversão da reação. No entanto, quando se adiciona amônio a uma reação em quantidades basais, acredita-se que haja uma estabilização estrutural ou então uma inibição direta pela interação do amônio com a enzima evitando a liberação dos produtos. O teste do efeito do amônio sobre a enzima (Figura 32) revelou que com relação à enzima sem adição alguma de amônio, há um aumento da atividade de PAL e HAL do agregado mPAL_c1 com concentrações ínfimas de amônio (até 0,1 mM). Não houve diferença significativa entre as atividades de PAL do agregado mPAL_c1 quando a concentração de amônio foi elevada em 10 vezes (de 0,1 a 1 mM), já a atividade de HAL foi reduzida linearmente até concentrações 100 vezes maiores. Em suma, o amônio sob concentração de 100 μ M pode ser um composto associado ao tampão de reação para melhorar a performance da mPAL_c1.

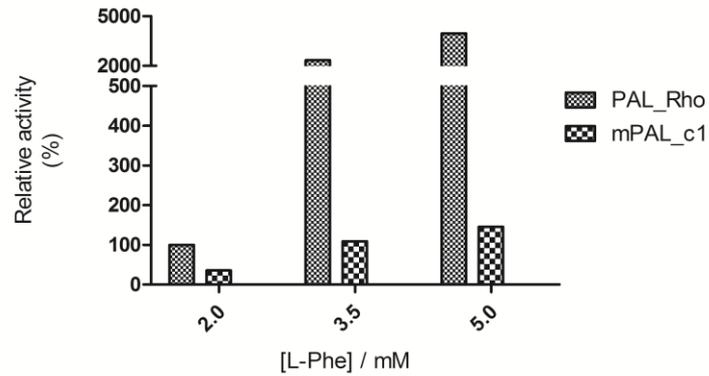


Figura 31. Ativação da mPAL_c1 por substrato e comparação com a enzima comercial de *R. glutinis*. A atividade enzimática mostrada é relativa, considerando-se 100% como a atividade de PAL_Rho a 2mM de L-Phe. A concentração de L-Phe variou de 2,0 mM a 5,0 mM, e mostrou-se capaz de ativar a PAL comercial quase 40 vezes, enquanto que a atividade do agregado mPAL_c1 inicial é muito mais baixa (perto de 30%) e sofre um aumento de apenas 5 vezes.

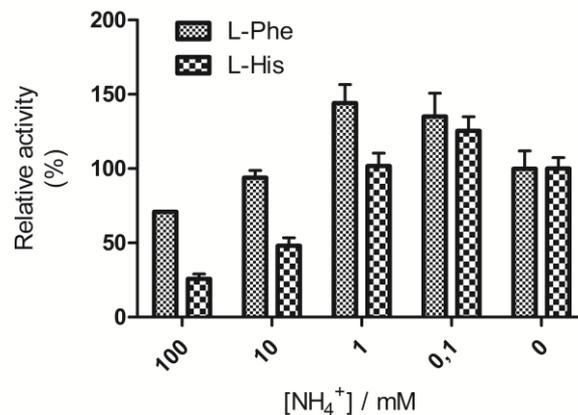


Figura 32. Efeito do amônio na mPAL_c1 com diferentes substratos. A atividade enzimática mostrada é relativa, considerando-se 100% como a atividade sem amônio, com cada substrato individualmente comparado. A concentração de NH_4^+ variou de 0 mM a 100 mM, e mostrou-se capaz de ativar a mPAL_c1 com ambos os substratos até concentrações de 0,1 mM.

CONCLUSÕES

- Teorias de que a recombinação teve papel fundamental na história evolutiva da PAL foram reforçadas pelo nosso estudo;
- O *design in silico* foi eficiente na geração de uma enzima quimérica e o protocolo determinado neste estudo é útil para novas aplicações deste tipo;
- O sistema de expressão heterólogo em *E. coli* não se revelou eficiente na produção de mPAL_c1;
- A purificação de mPAL_c1 envolveu uma abordagem de *refolding*, primeira vez relatado para enzimas do tipo PAL;
- A estrutura do agregado mPAL_c1 e dos dímeros e monômeros aparentes mostrou a forma esperada de um elipsóide oblongato, entretanto, a forma que possuiu atividade enzimática se revelou sem o enovelamento correto;
- O agregado mPAL_c1 apresentou atividade sobre os substratos comumente utilizados para PAL e HAL;
- A atividade enzimática do agregado mPAL_c1 foi ativada pelo substrato como previamente referido na literatura, e foi ativado também pelo íon amônio, quando em pequenas concentrações;
- A atividade da enzima mPAL_c1 é por volta de 1/3 da atividade da enzima comercial sob baixas concentrações de substrato, e por volta de 3% da atividade da enzima comercial quando em altas concentrações de L-Phe.

REFERÊNCIAS

- ABELL, C.W. *et al.*. The Effects of Phenylalanine Ammonia-Lyase on Leukemic Lymphocytes in Vitro. **Cancer Res.**, v. 32, p. 285-290, 1972.
- ACQUISTI, C.; ELSEER, J.J. e KUMAR S. Ecological nitrogen limitation shapes the DNA composition of plant genomes. **Mol. Biol. Evol.**, v. 26, p. 953–956, 2009.
- ACQUISTI, C.; KUMAR, S. e ELSEER, J.J. Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. **Proc. Biol. Sci.**, v. 276, p. 2605–2610, 2009.
- AKAIKE, H. A new look at the statistical model identification. *Autom. Control.* **IEEE Trans.**, v. 19, p. 716–723, 1974.
- AKASHI, H.; GOJOBORI, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. **Proc. Natl. Acad. Sci. U.S.A.**, v. 99, p. 3695–3700, 2002.
- ALUNNI, S.; CIPICIANI, A.; FIORONI, G. e OTTAVI, L. Mechanisms of inhibition of phenylalanine ammonia lyase by phenol inhibitors and phenol/glycine synergistic inhibitors. **Arch. Biochem. Biophys.**, v. 412, p. 170–175, 2003.
- AMBRUS, C.M. *et al.* Phenylalanine depletion for management of phenylketonuria: Use of enzyme reactors with immobilized enzymes. **Science**, v. 201, p. 837–839, 1978.
- AN, J.; ZHANG, M. e LU, Q. Changes in some quality indexes in fresh-cut green asparagus pretreated with aqueous ozone and subsequent modified atmosphere packaging. **J. Food Eng.**, v. 78, p. 340–344, 2007.
- ANSON, J.G. *et al.* Complete nucleotide sequence of the *Rhodospiridium toruloides* gene coding for phenylalanine ammonia-lyase. **Gene**, v. 58, p. 189-199, 1987.
- ARAÚJO, W.L. *et al.* (eds.) Guia Prático: Isolamento e caracterização de microrganismos endofíticos. 2 ed. Piracicaba: Copiadora Luiz de Queiroz, 2010.
- ASANO, Y. *et al.* Structure and Function of Amino Acid Ammonia-lyases. **Biocatalysis and Biotransformation**, v. 22, n.2, p. 131-138, 2004.
- AYRES, M.; AYRES JÚNIOR, M.; AYRES, D.L.; SANTOS, A.A. BIOESTAT – Aplicações estatísticas nas áreas das ciências biomédicas. 5ª ed. Belém, PA - Brazil, Ong Mamiraua, 2007.
- BAGAL, U.R. *et al.* Phylogenomic Analysis of the Phenylalanine Ammonia Lyase Gene Family in Loblolly Pine (*Pinus taeda* L.). **IEEE Intern. Conf. Bioinfo. Biom.**, p. 69-74, 2011.

- BAGAL, U.R. *et al.* The phenylalanine ammonia lyase (PAL) gene family shows a gymnosperm-specific lineage. **BMC Genomics**, v. 13, Suppl. 3: S1, 2012.
- BARRATT, E.S. *et al.* Effects of rapid depletion of phenylalanine and tyrosine on sleep and behavior. **Pharmacol. Biochem. Behav.**, v. 5, p. 47–53, 1976.
- BATTISTEL, E. e BIANCHI, D. Thermostability of ribonuclease in organic solvents: a calorimetric and spectroscopic study. **J. Phys. Chem.**, v. 98, p. 5368–5375, 1994.
- BEECHEM, J.M. e BRAND, L Time-resolved fluorescence of proteins. **Annu. Rev. Biochem.**, v. 54, p. 43–71, 1985.
- BJELLQVIST, B. *et al.* Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. **Electrophoresis**, v. 15, p. 529–539, 1994.
- BJELLQVIST, B. *et al.* The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. **Electrophoresis**, v. 14, p. 1023–1031, 1993.
- BIO-WORKS. Performance of WorkBeads™ 17/100 SEC preppacked size-exclusion column. Application n° 301. Inovata AB, Stockholm / Sweden. Acesso em: 15/01/2016. Disponível em: <http://www.biokal.com/assets/uploads/WorkBeads_17_SEC_AN301_on_performance.pdf>
- BÖHM, G. CDNN: CD spectra deconvolution software version 2.1. University of Halle-Wittenberg: Halle, 1997.
- BONI, M.F.; POSADA, D.; FELDMAN, M.W. An exact nonparametric method for inferring mosaic structure in sequence triplets. **Genetics**, v. 176, p. 1035–1047, 2007.
- BRAGG, J.G.; HYDER, C.L. Nitrogen versus carbon use in prokaryotic genomes and proteomes. **Proc. R. Soc. B.**, v. 271, p. S374–S377, 2004.
- BRAGG, J.G.; WAGNER, A. Protein material costs: single atoms can make an evolutionary difference. **Trends Genet.**, v. 25, p. 5–8, 2009.
- BRENDA-Enzyme. 2016. INFO: E.C. 4.3.1.24 - Phenylalanine Ammonia-lyase. In: Braunschweig Enzyme Database. Disponível em: <http://www.brenda-enzymes.info/php/result_flat.php4?ecno=4.3.1.24>. Acesso em: 02/01/2016.
- BRYANT, D.; MOULTON, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. **Mol. Biol. Evol.**, v. 21, p. 255–265, 2004.
- BURSTEIN, E.A. *et al.* Fluorescence and the location of tryptophan residues in protein molecules. **Photochem. Photobiol.**, v. 18, p. 263–279, 1973.

- BUTTERFOSS, G.L.; KUHLMAN, B. Computer-based design of novel protein structures. **Annu Rev Biophys Biomol Struct.**, v. 35, p. 49-65, 2006.
- CALABRESE, J.C. *et al.* Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. **Biochemistry**, v. 43, p. 11403–11416, 2004.
- CAMM, E.L. e TOWERS, G.H.N. Phenylalanine ammonia lyase. **Phytochemistry**, v. 12, p. 961–973, 1973.
- CAMPBELL, M.M. Elicited phenylpropanoid metabolism in pine cell cultures. [Dissertação de Mestrado]. Guelph: University of Guelph, 1991.
- CARLSON, R.P. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. **Bioinformatics**, v. 23, p. 1258–1264, 2007.
- CHAN, C.X. *et al.* Are protein domains modules of lateral genetic transfer? **PLoS One**, v. 4, e4524, 2009.
- CHEN, S. e GOPALAKRISHNAN, P. Speaker, environment and channel change detection and clustering via the Bayesian Information Criterion. In: *Proc. DARPA Broadcast News Transcription and Understanding Workshop*. p. 1–6, 1998.
- CHENG, C.K. e MARSH, H.V. Gibberellic Acid-Promoted Lignification and Phenylalanine Ammonia-lyase Activity in a Dwarf Pea (*Pisum sativum*). **Plant Physiol.**, v. 43, n. 11, p.1755-1759, 1968.
- CHESTERS, C. *et al.* Thermal Bifunctionality of Bacterial Phenylalanine Aminomutase and Ammonia Lyase Enzymes. **Angew Chem Int Ed**, v. 51, p. 4344–4348, 2012.
- CHRISTENSON, S. D. *et al.* Kinetic analysis of the 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis. **Biochemistry**, v. 42, p. 12708–12718, 2003.
- COCHRANE, F.C. *et al.* The *Arabidopsis* phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. **Phytochemistry**, v. 65, p. 1557–1564, 2004.
- COSCOLLÁ, M. *et al.* Genetic variability in environmental isolates of *Legionella pneumophila* from Comunidad Valenciana (Spain). **Environ. Microbiol.**, v. 8, p.1056–1063, 2006.
- COSTA, J. *et al.* Molecular evolution of key genes for type II secretion in *Legionella pneumophila*. **Environ. Microbiol.**, v. 14, p.2017–2033, 2012.
- COSTA, J. *et al.* 2010. Molecular evolution of *Legionella pneumophila* dotA gene, the contribution of natural environmental strains. **Environ. Microbiol.**, v. 12, p.2711–2729, 2010.

- COX, G.M. *et al.* The actin gene from *Cryptococcus neoformans*: structure and phylogenetic analysis. **J. Med. Vet. Mycol.**, v. 33, p. 261–266, 1995.
- CRAMER, C.L. *et al.* Phenylalanine ammonia-lyase gene organization and structure. **Plant Mol Biol**, v. 12, p.367-383, 1989.
- DAHIYAT, B.I.; MAYO, S.L. Protein design automation. **Protein Sci.**, v. 5, n. 5, p. 895-903, 1996.
- D’CUNHA, G.B. *et al.* Novel direct synthesis of L-phenylalanine methyl ester by using *Rhodotorula glutinis* phenylalanine ammonia lyase in an organic-aqueous biphasic system. **Enzyme Microb. Technol.**, v. 16, p. 318–322, 1994.
- D’CUNHA, G.B. *et al.* 1996. Stabilization of phenylalanine ammonia lyase containing *Rhodotorula glutinis* cells for the continuous synthesis of L-phenylalanine methyl-ester. **Enzyme Microb. Technol.**, v. 19, p. 421–427, 1996.
- DANIEL, R.M. *et al.* The effect of temperature on enzyme activity: new insights and their implications. **Extremophiles**, v. 12, p. 51–59, 2008.
- DAVIES, M.T. A universal buffer solution for use in ultraviolet spectrophotometry. **Analyst**, v. 84, p. 248-251, 1959.
- DAVIES, M.E. Effects of auxin on polyphenol accumulation and the development of phenylalanine ammonia-lyase activity in darkgrown suspension cultures of Paul's Scarlet rose. **Planta**, v. 104, n. 1, p. 66-77, 1972.
- DEAN, R.A. *et al.* 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. **Nature**, v. 434, p.980–986, 2005.
- DELPORT, W. *et al.* Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. **Bioinformatics**, v. 26, p. 2455–2457, 2010.
- DEMCHENKO, A.P. Ultraviolet Spectroscopy of Proteins. Springer-Verlag, New York, 1986.
- DEREEPER, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. **Nucleic Acids Res.**, v. 1, n. 36, p. W465-469, 2008.
- DING, F. *et al.* Mechanism for the α -Helix to β -Hairpin Transition. **PROTEINS: Structure, Function, and Genetics**, v. 53, p.220–228, 2003.
- DOBSON, C.M. Experimental investigation of protein folding and misfolding, **Methods**, v. 34, p. 4–14, 2004.
- DONLON, J. *et al.* Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: *The metabolic and molecular bases of inherited disease*. SCRIVER, C.R.; BEAUDET, A.R.; SLY, W. e VALLE, D. (Eds). Editora McGraw Hill, New York. pp. 1821–1838, 2004.

- DORON-FAIGENBOIM, A. *et al.* Selecton: a server for detecting evolutionary forces at a single amino-acid site. **Bioinformatics**, v. 21, p.2101–2103, 2005.
- EDGAR, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucl. Acids Res.**, v. 32, n. 5, p. 1792-1797, 2004.
- EDWARDS, K. *et al.* Rapid transient induction of phenylalanine ammonia-lyase mRNA in elicitor-treated bean cells. **Proc. Natl. Acad. Sci. U.S.A.**, v. 82, p.6731-6735, 1985.
- EFTINK, M.R. Fluorescence techniques for studying protein structure. **Methods Biochem. Anal.**, v. 35, p.127–205, 1991.
- ELSER, J.J. *et al.* Stoichiogenomics, p. the evolutionary ecology of macromolecular elemental composition. **Trends in Ecology and Evolution**, v. 26, n. 1, p. 38–44, 2011.
- EMILIANI, G. *et al.* A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. **Biology Direct**, v. 4, 7 (pp. 1-12), 2009.
- ENGELSMA, G. Low-temperature dependent development of phenylalanine ammonia-lyase in gherkin hypocotyls. **Naturwissenschaften**, v. 56, n. 11, p. 563, 1969.
- ENGELSMA, G.. Photoinduction of phenylalanine deaminase in gherkin seedlings: IV - The role of the temperature. **Planta**, v. 90, n. 2, p. 133-41, 1970a.
- ENGELSMA, G.. Low-temperature effects on phenylalnine ammonia-lyase activity in gherkin seedlings. **Planta**, v. 91, n. 3, p. 246-54, 1970b
- ERICKSON, H.P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. LI, S. (ed.), **Biological Procedures Online**, v. 11, n. 1, p. 32-51, 2009.
- ESTABROOK, E.M. e SENGUPTA-GOPALAN, C. Differential expression of phenylalanine ammonia-lyase and chalcone synthase during soybean nodule development. **Plant Cell**, v. 3, p. 299–308, 1991.
- EVANS, C. *et al.* Production of phenylalanine ammonia-lyase (PAL): isolation and evaluation of yeast strains suitable for commercial production of L-phenylalanine. **Appl. Microbiol. Biotechnol.**, v. 25, 1987a. DOI: 10.1007/BF00253309
- EVANS, C. *et al.* Novel stabilization of phenylalanine ammonia lyase catalyst during bioconversion of trans-cinnamic acid to L-phenylalanine. **Appl. Microbiol. Biotechnol.**, v. 25, p. 399–405, 1987b.
- EWING, B. e GREEN, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, p. 186-194, 1998.
- EWING, B. *et al.* Basecalling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v. 8, p. 175-185, 1998.

- FERRER, J.L. *et al.* Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. **Plant Physiol Biochem**, v. 46, p. 356-370, 2008.
- FILPULA, D. *et al.* Nucleotide sequence of gene for phenylalanine ammonia-lyase from *Rhodotorula rubra*. **Nucleic Acids Res.**, v. 16, n. 23, p. 11381, 1988.
- FINKELMAN, M.A.J. e YANG, H.H. Production of phenylalanine ammonia lyase by fermentation. **Genex Corporation**, European Pat.140714, 1985.
- FRANK, H.S. e FRANKS, F. Structural Approach to the Solvent Power of Water for Hydrocarbons; Urea as a Structure Breaker. **J. Chem. Phys.**, v. 48, p. 4746–4757, 1968.
- FRANK, R.L. e VODKIN, L.O. Sequence and structure of a phenylalanine ammonia-lyase gene from *Glycine max.* **J. DNA Sequencing Mapping**, v. 1, p. 335-346, 1991.
- FRITZ, R.R. *et al.* Phenylalanine ammonia lyase: induction and purification from yeast and clearance in mammals. **J. Biol. Chem.**, v. 251, p. 4646–4650, 1976.
- GÁMEZ, A. *et al.* Toward PKU enzyme replacement therapy: PEGylation with activity retention for three forms of recombinant phenylalanine hydroxylase. **Mol. Ther.**, v. 9, p. 124–129, 2004.
- GÁMEZ, A. *et al.* Development of Pegylated Forms of Recombinant *Rhodospiridium toruloides* Phenylalanine Ammonia-Lyase for the Treatment of Classical Phenylketonuria. **Mol. Ther.**, v. 11, n. 6, p. 986-989, 2005
- GARETH REES, D. e JONES, D.H. Stability of L-phenylalanine ammonia-lyase in aqueous solution and as the solid state in air and organic solvents. **Enzyme Microb. Technol.**, v. 19, p. 282–288, 1996.
- GARETH REES, D. e JONES, D.H. Activity of L-phenylalanine ammonia lyase in organic solvents. **Biochim. Biophys. Acta**, v. 1338, p. 121–126, 1997.
- GHOLIZADEH, A. e KOHNEHROUZ, B.B. Activation of phenylalanine ammonia lyase as a key component of the antioxidative system of salt-challenged maize leaves. **Braz. Soc. of Plant Physiol.**, v. 22, n. 4, p. 217-223, 2010.
- GIBBS, M.J. *et al.* Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. **Bioinformatics**, v. 16, p.573–582, 2000.
- GLOGE, A. *et al.* The behavior of substrate analogues and secondary deuterium isotope effects in the phenylalanine ammonia-lyase reaction. **Arch. Biochem. Biophys.**, v. 359, p. 1–7, 1998.
- GOLDSON, A. *et al.* Screening of phenylalanine ammonia lyase in plant tissues, and retention of activity during dehydration. **J. Sci. Food Agric.**, v. 88, p. 619-625, 2008.

- GOLDSOON-BARNABY, A. e SCAMAN, C. Purification and characterization of phenylalanine ammonia lyase from *Trichosporon cutaneum*. **Enzyme Research**, v. 2013, ID 670702, 2013.
- GORDON, D. e GREEN, P. Consed: A Graphical Editor for Next-Generation Sequencing. **Bioinformatics**, v. 29, n. 22 p. 2936-2937, 2013.
- GORDON, D. *et al.* Consed: A Graphical Tool for Sequence Finishing. **Genome Research**, v. 8, p.195-202, 1998.
- GORDON, D. *et al.* Automated Finishing with Autofinish. **Genome Research**, v. 11, n. 4, p.614-625, 2001.
- GOWRI, G. *et al.* Stress responses in alfalfa (*Medicago sativa*): 12 - Sequence analysis of phenylalanine ammonia-lyase (PAL) cDNA clones and appearance of PAL transcripts in elicitor-treated cell cultures and developing plants. **Plant Mol Biol**, v. 17, p. 415-429, 1991.
- GUINDON, S. e GASCUEL, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Syst. Biol.**, v. 52, p. 696–704, 2003.
- GURUPRASAD, K. *et al.* Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. **Protein Eng.**, v. 4, p.155–161, 1990.
- HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symp. Ser.**, v. 41, p.95–98, 1999.
- HAMBERGER, B. *et al.* Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. **Canadian Journal of Botany**, v. 85, n. 12, p. 1182-1201, 2007.
- HAMILTON, B.K. *et al.* Manufacture of L-amino acids with bioreactors. **Trends Biotechnol.**, v. 3, p. 64–68, 1985.
- HANSON, K.R. e HAVIR, E.A. L-phenylalanine ammonia lyase. II. Mechanism and kinetic properties of the enzyme from potato tubers. **Biochemistry**, v. 7, p. 1904–1914, 1968.
- HANSON, K.R. e HAVIR, E.A. L-phenylalanine ammonia lyase. IV. Evidence that the prosthetic group contains a dehydroalanyl residue and mechanism of action. **Arch. Biochem. Biophys.**, v. 141, p. 1–17, 1970.
- HANSON, K.R. *et al.* Stereochemical course of the elimination catalyzed by L-phenylalanine ammonia-lyase and the configuration of 2-benzamidocinnamic azlactone. **Chem. Comm.**, p. 185–186, 1971.

- HELGASON, T. *et al.* Phylogeny of the *Glomerales* and *Diversisporales* (Fungi: Glomeromycota) from actin and elongation factor 1-alpha sequences. **FEMS Microbiol. Lett.**, v. 229, p. 127–132, 2003.
- HODGINS, D.S. Yeast phenylalanine ammonia lyase: Purification, properties, and the identification of catalytically essential dehydroalanine. **J. Biol. Chem.**, v. 246, p. 2977–2985, 1971.
- HOLDER, M. e LEWIS, P.O. Phylogeny estimation: traditional and bayesian approaches. **Nature Reviews Genetics**, v. 4, p. 275 - 284, 2003.
- HOPKINS, W.G.; ORKWISZEWSKI, J.A.J. Inhibition of phenylalanine ammonia lyase and enhancement of *Avena* coleoptile segment elongation by fluorophenylalanine. **Canadian Journal of Botany**, v. 49, n. 1, p. 129-135, 1971.
- HSIA, D.Y. AND HOLTZMANN, N.A. A critical evaluation of phenylketonuria screening. In *Medical genetics*. McKu-SICK, V.A. e CLAIBORNE, R. (Eds.). HP Publishing, N.Y. pp. 237–244, 1973.
- HUSON, D.H. e BRYANT, D. Application of phylogenetic networks in evolutionary studies. **Mol. Biol. Evol.**, v. 23, p. 254–267, 2006.
- HYUN, M.W. *et al.* Fungal and Plant Phenylalanine Ammonia-lyase. **Mycobiology**, v. 39, n. 4, p. 257-265, 2011.
- INNERARITY, L.T. *et al.* Indoleacetic acid inhibition of a phenylalanine ammonia-lyase preparation from suspension cultures of WR-132 tobacco. **Phytochemistry**, v. 11, n. 1, p. 83-88, 1972.
- JEMTH, P. *et al.* Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. **Proc. Natl. Acad. Sci. U.S.A.**, v. 101, p. 6450– 6455, 2004.
- JIANG, L. *et al.* De novo computational design of retro-aldol enzymes. **Science**, v. 319, n. 5868, p. 1387-1391, 2008.
- JOOS, H.J. e HAHLBROCK, K. Phenylalanine ammonia-lyase in potato (*Solanum tuberosum* L.): Genomic complexity, structural comparison of 2 selected genes and modes of expression. **Eur. J. Biochem.**, v. 204, p.621-629, 1992.
- KAITISHA, G.C. Some virus diseases of crop plants in Zambia. In: *Plant virology in sub-Saharan Africa*. HUGHES, J. d’A. e ODU B.O. (Eds.). IITA, Nigeria, pp. 317–333, 2003.
- KÄMPER, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. **Nature**, v. 444, p. 97–101, 2006.

- KATSUYAMA, Y. *et al.* Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. **J. Biol. Chem.**, v. 284, n. 17, p. 11160-11170, 2009.
- KELLY, S.M. *et al.* How to study proteins by circular dichroism. **Biochimica et Biophysica Acta**, v. 1751, p. 119-139, 2005.
- KHAN, N.U. *et al.* Phenylalanine ammonia-lyase: An update of its kinetics. **Curr. Sci.**, v. 58, p. 427-430, 1989.
- KIM, S.H. *et al.* Cloning and disruption of a phenylalanine ammonia-lyase gene from *Ustilago maydis*. **Curr. Genet.**, v. 40, p. 40-48, 2001.
- KIM, S.H. *et al.* Purification and characterization of phenylalanine ammonia-lyase from *Ustilago maydis*. **Phytochemistry**, v. 43, p. 351-357, 1996.
- KLOSTERMAN, S.J. *et al.* Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. **PLoS Pathog.**, v. 7, e1002137, 2011.
- KONEV, S.V. Fluorescence and Phosphorescence of Proteins and Nucleic Acids. Plenum, New York, 1967.
- KOSAKOVSKY POND, S.L. *et al.* GARD: a genetic algorithm for recombination detection. **Bioinformatics**, v. 22, p. 3096-3098, 2006.
- KOUKOL, J. e CONN, E.E. Metabolism of aromatic compounds in higher plants. IV. Purification and properties of phenylalanine deaminase of *Hordeum vulgare*. **J. Biol. Chem.**, v. 236, p. 2692-2698, 1961.
- KOVÁCS, K. *et al.* Expression and Properties of the Highly Alkalophilic Phenylalanine Ammonia-Lyase of Thermophilic *Rubrobacter xylanophilus*. **PLoS ONE**, v. 9, n. 1, e85943, 2014.
- KRITTANAI, C. e JOHNSON JR., W.C. Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. **Anal. Biochem.**, v. 253, p. 57- 64, 1997.
- KUHN, D.N. *et al.* Induction of phenylalanine ammonia-lyase and 4-coumarate: CoA ligase mRNAs in cultured plant cells by UV light or fungal elicitors. **Proc. Natl. Acad. Sci. U.S.A.**, v. 81, p.1102-1106, 1984.
- KUMAR, A. e ELLIS, B.E.. The Phenylalanine Ammonia-Lyase gene family in Raspberry: Structure, expression, and evolution. **Plant Physiol.**, v. 127, p. 230-239, 2001.
- LAHR, D.J.G. *et al.* Comprehensive phylogenetic reconstruction of amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. **PLoS One**, v. 6, e22780, 2011.

- LANGER, B. *et al.* Methylidene-imidazolone (MIO) from histidine and phenylalanine ammonia lyase. **Adv. Protein Chem.**, v. 58, p. 175–214, 2001.
- LE, S.Q. e GASCUEL, O. An improved general amino acid replacement matrix. **Mol. Biol. Evol.**, v. 25, p.1307–1320, 2008.
- LEE, S.W. *et al.* Truncated phenylalanine ammonia-lyase expression in tomato (*Lycopersicon esculentum*). **J. Biol. Chem.**, v. 267, p.11824-11830, 1992.
- LIU, R. *et al.* Expression profile of a PAL gene from *Astragalus membranaceus* var. *Mongholicus* and its crucial role in flux into flavonoid biosynthesis. **Plant Cell Rep.**, v. 25, n. 7, p. 705-710, 2006.
- LOHMAN, J.R.; SHEN, B. Chapter 15: 4-Methylideneimidazole-5-one-containing aminomutases in enediyne biosynthesis. In: *Natural Product Biosynthesis by Microorganisms and Plants, Part B. Methods Enzymol.* v. 516, p. 299-319, 2012.
- LOIS, R. *et al.* A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. **EMBO J.**, v. 8, p.1641-1648, 1989.
- LÓPEZ-GALVEZ, G. *et al.* Wound induced phenylalanine ammonia lyase activity: Factors affecting its induction and correlation with the quality of minimally processed lettuce. **Postharvest Biol. Technol.**, v. 9, p. 223–233, 1996.
- LÓPEZ-VILLALOBOS, A. *et al.* Preservation of high phenylalanine ammonia lyase activities in roots of Japanese Striped corn: A potential oral therapeutic to treat phenylketonuria. **Cryobiology**, v. 68, n. 3, p. 436-445, 2014.
- LUIZ, D.P.; SANTOS JÚNIOR, C.D. *et al.* Tollip or not Tollip: What are the evolving questions behind it? **PLoS ONE**, v. 9, n. 5, ID e97219, 2014.
- MACDONALD, M.J. e CUNHA, G.B. A modern view of phenylalanine ammonia lyase. **Biochem. Cell Biol.**, v. 85, p. 273–282, 2007.
- MACDONALD, J.T. *et al.* Computational design approaches and tools for synthetic biology. **Integr. Biol.**, v. 3, p. 97–108, 2011.
- MADDISON, D. *et al.* The Tree of Life Web Project. **Zootaxa**, v. 1668, p.19–40, 2007.
- MARTIN, D. e RYBICKI, E. RDP: detection of recombination amongst aligned sequences. **Bioinformatics**, v. 16, p.562–563, 2000.
- MARTIN, D.P. *et al.* RDP3: a flexible and fast computer program for analyzing recombination. **Bioinformatics**, v. 26, p. 2462–2463, 2010.
- MARTIN, D.P. *et al.* A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. **AIDS Res. Hum. Retroviruses**, v. 21, p. 98–102, 2005.

- MARTIN, D.P. Recombination detection and analysis using RDP3. **Methods Mol. Biol.**, v. 537, p.185–205, 2009.
- MARUSIC, W.C. *et al.* Induction of L-phenylalanine ammonia lyase during utilization of phenylalanine as a carbon or nitrogen source in *Rhodotorula glutinis*. **J. Bacteriol.**, v. 146, p. 1013–1021, 1981.
- MASHIACH-FARKASH, E. *et al.* SymmRef: A flexible refinement method for symmetric multimers. **PROTEINS**, v. 79, n. 9, p. 2607-2623, 2011.
- MCGRATH, B.M. e WALSH, G. (Eds.) Directory of Therapeutic Enzymes. CRC Press: Boca Raton, 2006.
- MILNE, I. *et al.* TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. **Bioinformatics**, v. 25, p. 126–127, 2009.
- MINAMI, E. *et al.* Structure and some characterization of the gene for phenylalanine ammonia-lyase from rice plants. **Eur. J. Biochem.**, v. 185, p.19-25, 1989.
- MO, E.K. e SUNG, C.K. The effect of phenylethyl alcohol on PAL mRNA and enzyme activity in strawberries. **Postharvest Biol. Technol.**, v. 42, n. 3, p. 290-292, 2006.
- MOFFITT, M.C. *et al.* Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization. **Biochemistry**, v. 46, p.1004–1012, 2007.
- MOORE, B.S. *et al.* Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. **J. Nat. Prod.**, v. 65, n. 12, 1956-1962, 2002.
- MOORE, K. *et al.* Degradation of phenylalanine and tyrosine by basidiomycetes. **Life Sci.**, v. 6, p. 2629–2633, 1967.
- MORETTI, S. *et al.* PROTOGENE: turning amino acid alignments into bona fide CDS nucleotide alignments. **Nucleic Acids Res.**, v. 34, p.W600–603, 2006.
- MOZHAEV, V.V. *et al.* Homogenous solutions of hydrophilic enzymes in nonpolar organic solvents: new systems for fundamental studies and biocatalytic transformations. **FEBS Lett.**, v. 292, p. 159–161, 1991.
- NAKAMICHI, K. *et al.* Induction and stabilization of L-phenylalanine ammonia-lyase activity in *Rhodotorula glutinis*. **Appl. Microbiol. Biotechnol.**, v. 18, p. 158– 162, 1983.
- NEWTON, M.S. *et al.* Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. **J. R. Soc. Interface**, v. 12, n. 107, ID 20150036, 2015.
- NINIO, M. *et al.* Phylogeny reconstruction: increasing the accuracy of pairwise distance estimation using Bayesian inference of evolutionary rates. **ECCB**, v. 23, p. e136–e141, 2006.

- O'BRIEN, E.P. *et al.* Interactions between Hydrophobic and Ionic Solutes in Aqueous Guanidinium Chloride and Urea Solutions: Lessons for Protein Denaturation Mechanism. **J. Am. Chem. Soc.**, v. 129, p. 7346–7353, 2007.
- OHL, S. *et al.* Functional properties of a phenylalanine ammonia-lyase promoter from *Arabidopsis*. **Plant Cell**, v. 2, p. 837-848, 1990.
- OKADA, T. *et al.* Molecular characterization of the phenylalanine ammonia-lyase from *Ephedra sinica*. **Biol. Pharm. Bull**, v. 31, p. 2194-2199, 2008.
- OKADA, Y. e OKADA, M.J. Scavenging Effect of Water Soluble Proteins in Broad Beans on Free Radicals and Active Oxygen Species. **J. Agri. Food Chem.**, v. 46, p. 401–406, 1998.
- ORNDORFF, S.A. *et al.* Strain improvement of *Rhodotorula graminis* for production of a novel L-phenylalanine ammonia lyase. **Appl. Environ. Microbiol.**, v. 54, p. 996–1002, 1988.
- PACE, N.C. *et al.* How to measure and predict the molar absorption coefficient of a protein. **Prot. Sci.**, v. 4, p. 2411-2423, 1995.
- PADIDAM, M. *et al.* Possible emergence of new geminiviruses by frequent recombination. **Virology**, v. 265, p. 218–225, 1999.
- PEDROSA, F.O. *et al.* Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. **PLoS Genet.**, v. 7, e1002064, 2011.
- PILBÁK, S. *et al.* The essential tyrosine-containing loop conformation and the role of the C-terminal multi-helix region in eukaryotic phenylalanine ammonia-lyases. **FEBS J.**, v. 273, n. 5, p. 1004–1019, 2006.
- PLEISS, J. Protein design in metabolic engineering and synthetic biology. **Current Opinion in Biotechnol.**, v. 22, p. 611-617, 2011.
- POPPE, L. Methylidene-imidazolone: a novel electrophile for substrate activation. **Current Opinion in Chemical Biology**, v. 5, p. 512–524, 2001.
- POSADA, D. e BUCKLEY, T.R. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. **Syst. Biol.**, v. 53, p. 793–808, 2004.
- POSADA, D. e CRANDALL, K.A. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. **Proc. Natl. Acad. Sci. U.S.A.**, v. 98, p. 13757–13762, 2001.
- POSADA, D. jModelTest: phylogenetic model averaging. **Mol. Biol. Evol.**, v. 25, p. 1253–1256, 2008.

- RADFORD, S.E. Protein folding: progress made and promises ahead. **Trends Biochem. Sci.**, v. 25, p. 611 – 618, 2000.
- RESHETNYAK, Y.A.K. e BURSTEIN, E.A. Assignment of the components of the fluorescence spectrum of protein to tryptophan residues based on the properties of their microenvirons in a three dimensional structure. **Biophysics**, v. 42, p. 267–274, 1997.
- RITTER H., SCHULZ G.E. Structural Basis for the Entrance into the Phenylpropanoid Metabolism Catalyzed by Phenylalanine Ammonia-Lyase. **Plant Cell**, v. 16, p. 3426–3436, 2004.
- ROHL, C.A. *et al.* Protein structure prediction using Rosetta. **Methods Enzymol.**, v. 383, p. 66-93, 2004.
- ROKAS, A. The effect of domestication on the fungal proteome. **Trends Genet.**, v. 25, p. 60–63, 2009.
- ROTHER, D. *et al.* An active site homology model of phenylalanine ammonia lyase from *Petroselinum crispum*. **Eur. J. Biochem.**, v. 269, p. 3065–3075, 2002.
- ROTHER, D. *et al.* Characterization of the active site of histidine ammonia-lyase from *Pseudomonas putida*. **Eur. J. Biochem.**, v. 268, p. 6011-6019, 2001.
- RÖTHER, D. *et al.* Spectroscopic Evidence for a 4-Methylidene Imidazol-5-one in Histidine and Phenylalanine Ammonia-Lyases. **Angew. Chem. Int. Ed.**, v. 39, p. 2462–2464, 2000.
- ROTHLISBERGER, D. *et al.* Kemp elimination catalysts by computational enzyme design. **Nature**, v. 453, n. 7192, p. 190-195, 2008.
- SAITOU, N. e IMANISHI, T. Relative efficiencies of the Fitch-Margoliash maximum-evolution, maximum likelihood, minimum evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. **Mol. Biol. Evol.**, v. 6, p. 514-525, 1989.
- SAITOU, N. e NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.**, v. 4, p. 406-425, 1987.
- SAMBROOK, J. e RUSSEL, D.W. *Molecular Cloning: A Laboratory Manual* (3a ed., 2028 p.). New York, USA: Cold Spring Harbor Laboratory, 2001.
- SARKISSIAN, C.N. e GÁMEZ, A. Phenylalanine ammonia lyase, enzyme substitution therapy for phenylketonuria, where are we now? **Mol. Genet. Metab.**, v. 86, Suppl 1, p. S22–26, 2005.
- SATO, T. *et al.* Inhibition of phenylalanine ammonia lyase by cinnamic acid derivatives and related compounds. **Phytochemistry**, v. 21, p. 845–850, 1982.

SCHÜRMAN, G. *et al.* Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. **Mol Biol Cell**, v. 12, p.1765–73, 2001.

SCHUSTER, B. e RETEY, J. The mechanism of action of phenylalanine ammonia lyase: the role of prosthetic dehydroalanine. **Proc. Natl. Acad. Sci. U.S.A.**, v. 92, p. 8433–8437, 1995.

SCHWEDE, T.F. *et al.* Crystal structure of histidine ammonia lyase revealing a novel polypeptide modification as the catalytic electrophile. **Biochemistry**, v. 38, p. 5355–5361, 1999.

SCRIVER, C.R. e KAUFMAN, S. Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: *The metabolic and molecular bases of inherited disease*. McGRAW HILL, New York, 2001.

SEDEROFF, R. *et al.* Genetic regulation of lignin biosynthesis and the potential modification of wood by genetic engineering in loblolly pine. **Recent. Adv. Phytochem.**, v. 28, p. 313-355, 1994.

SHARPTON, T.J. *et al.* Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. **Genome Res.**, v. 19, p. 1722–1731, 2009.

SHAW, N.M. *et al.* Wound induced phenylalanine ammonia-lyase in potato (*Solanum tuberosum*) tuber discs. Significance of glycosylation and immunolocalization of enzyme subunits. **Biochem. J.**, v. 267, p. 163–170, 1990.

SIEGEL, L.M. e MONTE, K.J. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. **Biochim Biophys Acta**, v. 112, p.346–362, 1966.

SIEGEL, J.B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. **Science**, v. 329, n. 5989, p. 309-313, 2010.

SIN, F.Y.T. *et al.* Molecular Evolution and Phylogeny of Actin Genes in *Haliothis* Species (Mollusca Gastropoda). **Zool. Stud.**, v. 46, p. 734–745, 2007.

SINGH, S.M. e PANDA, A.K. Solubilization and refolding of bacterial inclusion body proteins. **Journal of Bioscience and Bioengineering**, v. 99, n. 4, p. 303-310, 2005.

SMITH, J.M. Analyzing the mosaic structure of genes. **J. Mol. Evol.**, v. 34, p.126–129, 1992.

STAFYLAKIS, T. *et al.* Redefining the Bayesian Information Criterion for Speaker Diarisation. In: p. INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association. Brighton, United Kingdom. p. 1051–1054, 2009.

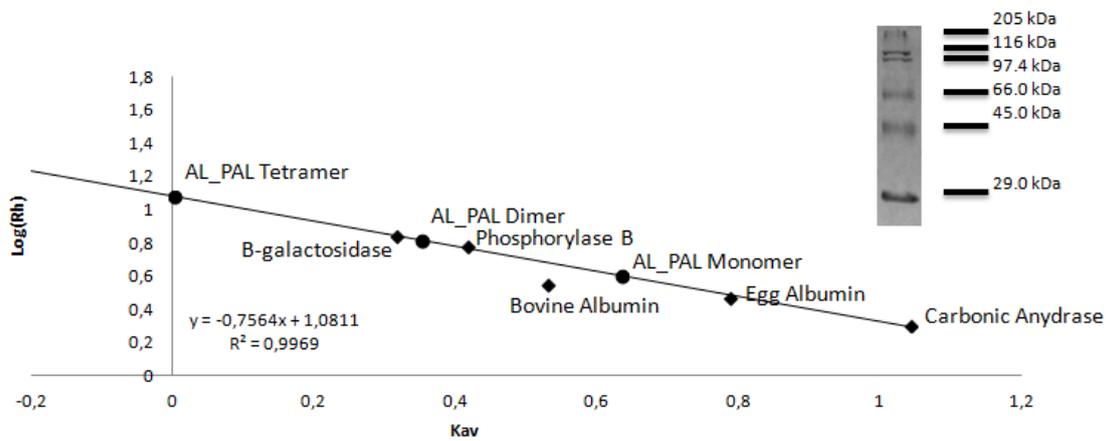
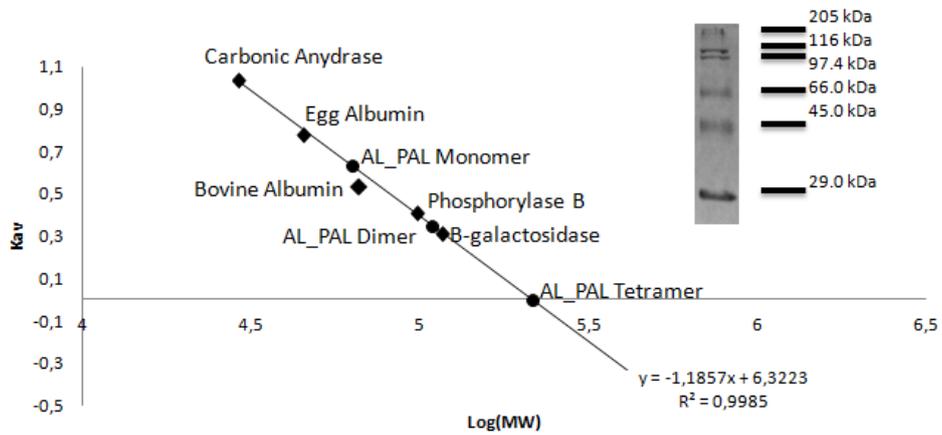
STAJICH, J.E. *et al.* Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). **Proc. Natl. Acad. Sci. U.S.A.**, v. 107, p.11889–11894, 2010.

- STERN, A. *et al.* Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. **Nucleic Acids Res.**, v. 35, p. W506–511, 2007.
- SUBRAMANIAM, R. *et al.* Structure, inheritance, and expression of hybrid poplar (*Populus trichocarpa* x *Populus deltoides*) phenylalanine ammonia-lyase genes. **Plant Physiol**, v. 102, p. 71-83, 1993.
- SWANSON, W.J. *et al.* Pervasive adaptive evolution in mammalian fertilization proteins. **Mol. Biol. Evol.**, v. 20, p. 18–20, 2003.
- TAIZ, L. e ZEIGER, E. *Plant Physiology*. Ohio, Massachusetts: Sinauer Associates, Inc. Publishers - 2006.
- TAMURA, K. *et al.* MEGA6, p. Molecular Evolutionary Genetics Analysis version 6.0. **Mol. Biol. Evol.**, v. 30, p. 2725–2729, 2013.
- TANAKA, Y. *et al.* Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase from cut-injured roots of sect potato. **Plant Physiol**, v. 90, p. 1403-1407, 1989.
- TAYLOR, R.G. *et al.* Cloning and Expression of Rat Histidase: homology to two bacterial histidases and four phenylalanine ammonia-lyases. **The Journal of Biological Chemistry**, v. 265, p. 18192-18199, 1990.
- TOBI, D. *et al.* The Dominant Interaction between Peptide and Urea Is Electrostatic in Nature: A Molecular Dynamics Simulation Study. **Biopolymers**, v. 68, p. 359–369, 2003.
- TSUMOTO, K. *et al.* Practical considerations in refolding proteins from inclusion bodies. **Protein Expression and Purification**, v. 28, p. 1-8, 2003.
- VASLET, C.A. *et al.* cDNA and genomic cloning of yeast phenylalanine ammonia-lyase reveal genomic intron deletions. **Nucleic Acids Res**, v. 16, p. 11382, 1988.
- VIVIAN, J.T. *et al.* Mechanisms of Tryptophan Fluorescence Shifts in Proteins. **Biophysical Journal**, v. 80, p. 2093–2109, 2001.
- WALTON, D.C. e SONDHEIMER, E. Effects of Abscisin II on Phenylalanine Ammonia-Lyase Activity in Excised Bean Axes. **Plant Physiol.**, v. 43, n. 3, p. 467-469, 1968.
- WANG, L. *et al.* Structural and biochemical characterization of the therapeutic *Anabaena variabilis* phenylalanine ammonia lyase. **J Mol Biol**, v. 380, p. 623–335, 2008.
- WANG, L. *et al.* Structure-based chemical modification strategy for enzyme replacement treatment of phenylketonuria. **Mol Genet Metab**, v. 86, p. 134–140, 2005.

- WANNER, L.A. *et al.* The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. **Plant Mol Biol.**, v. 27, n. 2, p. 327-338, 1995.
- WATANABE, S.K. *et al.* Phenylalanine ammonia lyase from *Sporidiobolus pararoseus* and *Rhodospiridium toruloides*: application for phenylalanine and tyrosine deamination. **World J. Microbiol. Biotechnol.**, v. 8, p. 406-410, 1992.
- WHETTEN, R.W. e SEDEROFF, R.R. Phenylalanine ammonia lyase from loblolly pine: Purification of the enzyme and isolation of complementary DNA clones. **Plant Physiol.**, v. 98, p. 380-386, 1992.
- WILKINS, M.R. *et al.* Protein identification and analysis tools in the ExPASy server. **Methods Mol. Biol.**, v. 112, p. 531-552, 1999.
- WILLIAMS, V.R. e HIROMS, J.M. Reversibility of the “irreversible” histidine ammonia-lyase reaction. **Biochim. Biophys. Acta**, v. 139, p. 214-216, 1967.
- XU, F. *et al.* Molecular cloning, characterization and expression of phenylalanine ammonia-lyase gene from *Ginkgo biloba*. **African J. of Biotech.**, v. 7, p. 721-729, 2008.
- YAMADA, S. *et al.* Production of L- phenylalanine from trans-cinnamic acid with *R. glutinis* containing L-phenylalanine ammonia lyase activity. **Appl. Environ. Microbiol.**, v. 42, p. 773-778, 1981.
- YANG, Z. *et al.* Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics**, v. 155, p. 431-449, 2000.
- YANG, Z. *et al.* Highly Efficient Production of Soluble Proteins from Insoluble Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. **PLoS ONE**, v. 6, n. 7, e22981, 2011.
- YANG, Z. Handbook of Statistical Genetics. New York, USA: Wiley, 2007. p. 377-406.
- ZAKS, A. e KLIBANOV, A.M. Enzymatic catalysis in organic media at 100 degrees C. **Science**, v. 224, p. 1249-1251, 1984.
- ZHANG, D. *et al.* The actin multigene family in *Populus*: organization, expression and phylogenetic analysis. **Mol. Genet. Genomics**, v. 284, p. 105-119, 2010.
- ZANGHELLINI, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. **Protein Sci.**, v. 15, n. 12, p. 2785-2794, 2006.
- ZIMMERMAN, A. e HAHLBROCK, K. Light induced changes in enzyme activities in parsley cell suspension. Purification and some properties of phenylalanine ammonia lyase. **Arch Biochem Biophys**, v. 166, p. 54-62, 1975.

ZUCKERKANDL, E. e PAULING, L. Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins*. BRYSON, P. e VOGEL, H. (Eds.). New York, USA, Academic Press, 1965. p. 97–166.

ANEXO III - Curvas de calibração para a cromatografia de exclusão molecular em coluna HiPrep Sephacryl S300 16/60 HR



ANEXO IV - Proteínas envolvidas com o sistema de transferência de core conservado

- Extremidade N-terminal

> LT159968 | 8LP_44363.118104_827
AANKITACRNYLDKLLDDES AVFYGINTGFGFLQNVQIDATQTEQLQYNLLMSHACGLGEQVPVEIVRLMLLLKIKSL
SYGN SGVQLQTVTRLIELFNHDVFPFIIFTQGS LGASGDLAPLSHLSLPLIGLGEVIHHQKKQLAKHVLQHFNWQPI SL
KSKEGLALINGTQFMSAYGMYCLYKASHLIK MADYISALS FDAFDCSTEALHPLIHKIRPHKGQVNTAKNILQYLQRS
EISQQKKKQIQDPYSFRCPVQVHGATKDTFEYVLNVFEKEIN

Domínio conservado (CDD): PRK09367

Descrição: Histidina Amônia Liase (Provisório)

Intervalo: 1-276

E-valor: 1.57e-121

Blastp Best-Hit

Histidine ammonia-lyase [*Sediminibacterium* sp. OR43] (WP_051348018.1)

E-valor = 1e-139 // Identidade = 70% // Cobertura = 100%

- Extremidade C-terminal

> LT159969 | 8LP_5286.33660_929
ALRGLFEAEDLF AAATVCGGLSVEAMLGSRAFPDARIHEVRGQRQIDVAAA YRDLLTDSSEIAHSHEDCSKVQDPYS
LRCQPQVMGACLTQIRQAAEVLEIEANAVSDNPLVFAEQGDVISGGNFHAEPVAMAADNLALATAEIGSLSERRISMM
MDRHMSQLPPFLVANGGVNSGFMIAQVTAALASDNKALAH PASVDSLPTSANQEDHVS MAPNAGKRLWYMADNVRGI
LAVEWLGACQGLDFREGLKSSPKLEQARKILRAQVPYYS EDRFFAPDIEQASELLSSGCLNELIIPKLLPSLSEV

Domínio conservado (CDD): PRK09367

Descrição: Histidina Amônia Liase (Provisório)

Intervalo: 1-298

E-valor: 0e+00

Blastp Best-Hit

MULTISPECIES: histidine ammonia-lyase [*Acinetobacter calcoaceticus/baumannii* complex] (WP_005804357.1)

E-valor = 0e+00 // Identidade = 100% // Cobertura = 100%

- Proteína Híbrida final mPAL_c1

> LT159967 | mPAL_c1
MGSSHHHHHSSGLVPRGSHMAANKITACRNYLDKLLDDES AVFYGINTGFGFLQNVQIDATQTEQLQYNLLMSHACG
LGEQVPHSATRAAMLVRINTLLQGYS GIRFEILEAITKLLNHNITPCLPLRGTISASGDLVPLSYIAGLLTGLGEVIH
HQKKQLAKHVLQHFNWQPFELQPK EGLALVNGTAVGSLASLVLFEANILAVLNASDEVLSANILNFS AENAVMQGNK
SPEFTDPLNITHKLNKRHHPGQIEAAAIMEHILTDSSEIAHSHEDCDGSQDNPRYNASLRNCTSPQNVWLGNAPQIEV
IRFATKNSVIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNNGLPSN
LVLNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTAEAVDILKMSATFLYALCQA
IDLRGEEIDSPKLEQARKVFTAQVPYYS EDRFFAPDIEQASELLSSGCLNELIIPKLLPSLSEV

Domínio conservado (CDD): TIGR01226 (phe_am_lyase)

Descrição: Fenilalanina Amônia Liase

Intervalo: 25-490

E-valor: 0e+00

Blastp Best-Hit

Phenylalanine ammonia-lyase [*Musa acuminata*]

E-valor = 0e+00 // Identidade = 72% // Cobertura = 80%

ANEXO V - Alinhamento final mPAL_c1 e as sequências parentais: O padrão da sequência final em *scramble*

```

8LP_5286.33660_929 .....
mPAL_c1 MGSSHHHHHHSSGLVPRGSHM AANKITACRNYLDKKLDDES AVFYGINTGFGFLQNVQID
8LP_44363.118104_827 ..... AANKITACRNYLDKKLDDES AVFYGINTGFGFLQNVQID
consensus>50 ..... aankitacrnyldkklddesavfygintgfgflqnvqid
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ATQTEQLQYNLLMSHACGLGEQVP HSATRAAMLVRIINTLLQGYSGIRF EILEAITKLLNH
8LP_44363.118104_827 ..... ATQTEQLQYNLLMSHACGLGEQVP VEIVRLMLLRLKLSYGN SGVQLQTVTRLIELFNH
consensus>50 ..... atqteqlqynllms hacglgeqvp . . . . r . ml . l . i . l . . g . sgi . . e . . . . l . nh
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 NITPCLPLRGTISASGDI VPLSYIAGLITGLGEVIHHQKKQLAKHVLQHFNWQPFELQPK
8LP_44363.118104_827 ..... DVFP I I I F T Q G S L G A S G D I A P L S H L S L P L I G L G E V I H H Q K K Q L A K H V L Q H F N W Q P I S L K S K
consensus>50 ..... ni . p . . . . . g . . . . . asgdl . pls . . . . . l . glgevi h h q k k q l a k h v l q h f n w q p . . l . . k
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 1 10 20 30
EGLALVNGTAVG SGLASLV LFEANI LAV LNASDEVLSANILNFSAENAVMQGNKSP EFTD
8LP_44363.118104_827 ..... EGLALVNGTAVG SGLASLV LFEANI LAV LNASDEVLSANILNFSAENAVMQGNKSP EFTD
consensus>50 ..... eglalvngt . . . . . s . . . . . lf . a . . l . . l . . a # dv . aa . . . . . f . . s . ea $ . g . . . . . p . f . .
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 40 50 60 70 80
R I H . . . . . E V R G Q R G Q I D V A A A Y R D L L T D S S E I A H S H E D C S K V Q D P . . . . . Y S I R C Q P Q V M
8LP_44363.118104_827 ..... R I H . . . . . E V R G Q R G Q I D V A A A Y R D L L T D S S E I A H S H E D C S K V Q D P . . . . . Y S I R C Q P Q V M
consensus>50 ..... . ih . . . . . ev R . h . G Q ! # . Aa . . . . . di l t d s s e i a h s h # d c . . v Q D p . . . . . y s I R c . p q v .
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 90 100 110 120 130
G A C L T Q I R C A A . . . . . E V L E I E A N A V S D N P L V F A E Q G D V I S G G N F H A E P V A M A A D N L
8LP_44363.118104_827 ..... G A T K D T F E Y V L . . . . . N V F E K E I N . . . . . D V S R N K A L H G G N F Q G T P I G V S M D N T
consensus>50 ..... ga . l . q . . qi . . . . . ev . E . E i N . v . d n p l v . . . . . g g n f . . . . . p v . . . . . d n .
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 140 150 160 170 180 190
A L A I A E I G S T S E R R I S M M M . D R H M S Q L P P F L V A N G G V N S G F M I A Q V T A A A L A S D N K A L A H
8LP_44363.118104_827 ..... R L A L A S I G K L M F A Q F S E L V N D F Y N N G L P S N L V L N P S L D Y G F K G A E I A M A S Y C S E L Q Y L A N
consensus>50 ..... . la . a . ig . l . . . . . s . m . d . . . . . lp . . lv . n . . . . . n . gf . . . . . aqv . . . . . sd . . . . . la .
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 200 210 220 230 240 250
P A S V D S L P T S A N Q E D H V S M A P N A G K R L W Y M A D N V R G I L A V E W L G A C Q G L D F R E G L K S S P K
8LP_44363.118104_827 ..... P V T S H V Q S A E Q H N Q D V N S L G L I S S R K T A E A V D I L K L M S A T F L Y A L C Q A I D L R G E E I D S P K
consensus>50 ..... p . . . . . qed . sm . . . . . d . . . . . a . . . . . cq . . . . . d r . . . . . spk
    
```

```

8LP_5286.33660_929 .....
mPAL_c1 ..... 260 270 280 290 300
L E Q A R K I L R A Q V P Y S E D R F F A P D I E Q A S E L L S S G C L N E L I I P K L L P S L S E V
8LP_44363.118104_827 ..... L E Q A R K V F T A Q V P Y S E D R F F A P D I E Q A S E L L S S G C L N E L I I P K L L P S L S E V
consensus>50 ..... leqark i . . . . . aqvpyysedrffapdieqasellssgclneliipkllpslsev
    
```

ANEXO VI - Alinhamento da mPAL_c1 e seqüências de PAL/HAL com a estrutura resolvida

1W27|*Petroselinum_cr* α1 α2 β1
 1 10 20 30 40 50
 1W27|*Petroselinum_cr* MENGNGATNGHVNNGMDFCMKTEDPLYWGIAAEAMTGSHLDEVKKMVAEYR..KPVVK
 4C5U|*Taxus_chinensis* ...MGSSHHHHHSSGL...VPRGSHMGFAVESR..SHVKDILGLINA FNEVKKITV
 4BAA|Redesign_1 ...MGSSHHHHHSSGL...VPRGSHMGFAVESR..SHVKDILGLINA FNEVKKITV
 2YII|Redesign_2 ...MGSSHHHHHSSGL...VPRGSHMGFAVESR..SHVKDILGLINA FNEVKKITV
 3NZ4|*Taxus* ...MGSSHHHHHSSGL...VPRGSHMGFAVESR..SHVKDILGLINT FNEVKKITV
 1Y2M|*Rhododporidium_mPAL_c1* XAPSLDSISHSFANGVASAKQAVNGASTNLVAVAGSHLPTTQVTVQV DIVEKXLAAPDSTL
 2NYF|*N_punctiforme*MNITSLQONITRSWQIPFTNSSDSI
 2NYN|*Anabaena*MKTLSQAQSKTSSQFSFTGNSANV
 1GK3|*HAL_Pseudomonas*T
 1GKJ|*HAL_Pseudomonas*T
 consensus>50a.....di.....f.....v

1W27|*Petroselinum_cr* α3 β2 η1 α4
 60 70 80 90 100 110
 1W27|*Petroselinum_cr* LGGETLTISQVAAASARDGSGVTVELEAARAGVKASDSDWVMDSMNKGTDIS.YGVTTGFG
 4C5U|*Taxus_chinensis* DGTTPITVAHVAAALARRHDDVVKVALE.AEQCRARVECTSSWVQRKAEEDGADI.YGVTTGFG
 4BAA|Redesign_1 DGTTPITVAHVAAALARRHDDVVKVALE.AEQCRARVECTSSWVQRKAEEDGADI.YGVTTGFG
 2YII|Redesign_2 DGTTPITVAHVAAALARRHDDVVKVALE.AEQCRARVECTSSWVQRKAEEDGADI.YGVTTGFG
 3NZ4|*Taxus* DGTTPITVAHVAAALARRHDDVVKVALE.AEQCRARVECTSSWVQRKAEEDGADI.YGVTTGFG
 1Y2M|*Rhododporidium_mPAL_c1* ELDGYSLNLGDDVVSARKGRPVVKDSDERTSKIDKSVFELRSQLS..XSV.YGVTTGFG
 2NYF|*N_punctiforme* ...MGSSHHHHHSSGLVPRGSHMAANKITACRNYLDLKKLLDDESAVF.YGVTTGFG
 2NYN|*Anabaena* TVGDRNLTIDEVNVVARHGTQVRLTDNADVIRGVOASCDYINNAVETAGQPI.YGVTTGFG
 1GK3|*HAL_Pseudomonas* IIGNQKLTINDVAVARNHTLVSLLTNNNDILQGIQASCDYINNAVETAGQPI.YGVTTGFG
 1GKJ|*HAL_Pseudomonas* ELTLKPGTLLAQLRAIHAAPVRLQLDASAAPADASVACVEQIIAEDRTA.YGINTGFG
 consensus>50 d.t.....aal.arh...v.le.eq.r..!eas...vq...edg...i.YGIttGFG

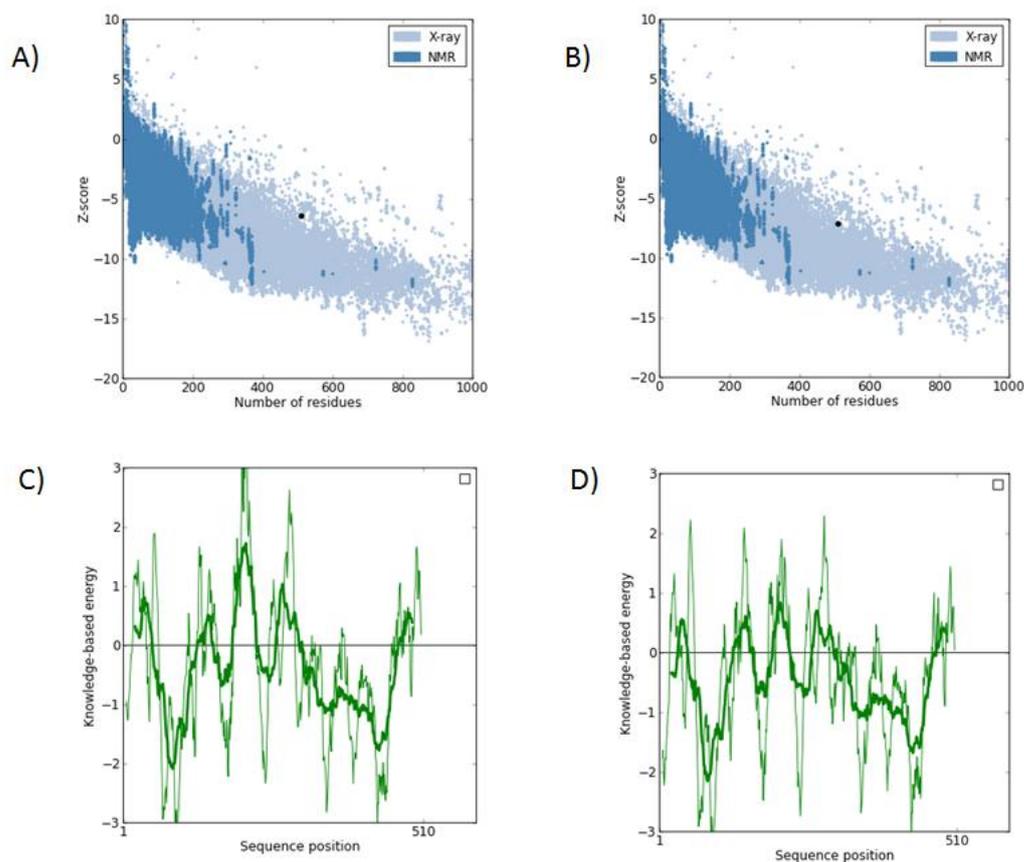
1W27|*Petroselinum_cr* α5 α6
 TT 120 130 140 150 160
 1W27|*Petroselinum_cr* ATSHRRTRK..QGALQKELEIRFTNAGIFGNG...S...DNTLPHSATRAAMLVRINT
 4C5U|*Taxus_chinensis* ACSRRRTN..RLSELEESLIRCLLAGVFTKGCAPS...VDELTPATATRSAMLRLNS
 4BAA|Redesign_1 ACSRRRTN..RLSELEESLIRCLLAGVFTKGCAPS...VDELTPATATRSAMLRLNS
 2YII|Redesign_2 ACSRRRTN..RLSELEESLIRCLLAGVFTKGCAPS...VDELTPATATRSAMLRLNS
 3NZ4|*Taxus* ACSRRRTN..RLSELEESLIRCLLAGVFTKGCASS...VDELTPATVTRSAMLRNS
 1Y2M|*Rhododporidium_mPAL_c1* GSADTRTE.DAISELQKALLEHQLCGVLPSSFDSFRLRGRLENSLPLVEVVRGAXTRVNS
 2NYF|*N_punctiforme* FLQNVQIDATQTEOQLQYNNLLMSHACGL...GEQVPHSATRAAMLVRINT
 2NYN|*Anabaena* GMADVVISREQAASELQTNLIWFLKSGA...GNKLSLADVRAAMLRLNS
 1GK3|*HAL_Pseudomonas* GMADVVISREQAASELQTNLIWFLKSGA...GNKLSLADVRAAMLRLNS
 1GKJ|*HAL_Pseudomonas* LLASTRIASHDLENLRSVLVSHAAGI...GAPLDDDLVRLIMVVKINS
 consensus>50 LLASTRIASHDLENLRSVLVSHAAGI...GAPLDDDLVRLIMVVKINS
 ...s.rtn..ql.eLQ.sLi..l.aGv.....n.lp...tR.amllr.Ns

1W27|*Petroselinum_cr* α7 β3 α8 β4
 170 180 190 200 210 220
 1W27|*Petroselinum_cr* LLQGYSCIRFEILEATKFLNNTIPCLPLRGTITX..DLVPLSYIAGLLIGRPNKAVG
 4C5U|*Taxus_chinensis* FTYGCSGIRFEVMEALEKLLNENVPKVLIRGSVSAAGDLPFLAYIAGLLIGKPSVIARI
 4BAA|Redesign_1 FTYGCSGIRFEVMEALEKLLNENVPKVLIRGSVSAAGDLPFLAYIAGLLIGKPSVIARI
 2YII|Redesign_2 FTYGCSGIRFEVMEALEKLLNENVPKVLIRGSVSAAGDLPFLAYIAGLLIGKPSVIARI
 3NZ4|*Taxus* FTYGCSGIRFEVMEALEKLLNENVPKVLIRGSVSAAGDLPFLAYIAGLLIGKPSVIARI
 1Y2M|*Rhododporidium_mPAL_c1* LTRGHSAVRFLVLEALTNFLNPGITPIVPLRGTISASGDLSPSYIAGLLIGHPDCKVHV
 2NYF|*N_punctiforme* LLQGYSCIRFEILEATKFLNNTIPCLPLRGTISASGDLSPSYIAGLLIGLGEVHH.
 2NYN|*Anabaena* HLYGASGIRLELIQRIETFLNAGVTPHYVEFGSIX..DLVPLSYITGALIGLDPSTFVD
 1GK3|*HAL_Pseudomonas* HMRGASGIRLELIKRMEIFLWAGVTPHYVEFGSIX..DLVPLSYITGALIGLDPSTFVD
 1GKJ|*HAL_Pseudomonas* LSRGFSGIRKVIDALIALVNAEVYPHIFPLKGSVGAAGALAPLAHMSLVLLGEGKARY..
 consensus>50 LSRGFSGIRKVIDALIALVNAEVYPHIFPLKGSVGAAGALAPLAHMSLVLLGEGKARY..
 ...G.Sg!R.evmealekllN.n!.P.vplrGs!sx..dLvPLayiaGliG.p.....

1W27|*Petroselinum_cr* β5 α9 α10 α11
 TT 230 240 250 260 270 280
 1W27|*Petroselinum_cr* ...PTGVTLSPEEAFKLAQVEGGEFELQPKKEGLALVNGTAVSGMASMVLFEANTLAVLA
 4C5U|*Taxus_chinensis* ...GDDVEVPAPEALSRVGLRP..FKLQAKEGLALVNGTSFATAVASTVMDANVLLLV
 4BAA|Redesign_1 ...GDDVEVPAPEALSRVGLRP..FKLQAKEGLALVNGTSFATAVASTVMDANVLLLV
 2YII|Redesign_2 ...GDDVEVPAPEALSRVGLRP..FKLQAKEGLALVNGTSFATAVASTVMDANVLLLV
 3NZ4|*Taxus* ...GDDVEVPAPEALSRVGLRP..FKLQAKEGLALVNGTSFATAVASTVMDANVLLLV
 1Y2M|*Rhododporidium_mPAL_c1* VHEGKKEKILYAREAXALFNLEP..VVLQPKKEGLGLVNGTAVSASXATLALDHAXLSSL
 2NYF|*N_punctiforme* ...QKKQLAKHVLOHFNWQP..FELQPKKEGLALVNGTAVSGLASLVLFEANTLAVLN
 2NYN|*Anabaena* ...FDGKEMDAPTALRQLNLPK..LQLQPKKEGLAMNGTSVMTGIAANCVYDQVLLALT
 1GK3|*HAL_Pseudomonas* ...FNGKEMDAPTALRQLNLPK..LQLQPKKEGLAMNGTSVMTGIAANCVYDQVLLALT
 1GKJ|*HAL_Pseudomonas* ...KQQLWLSATEALAVAGLEP..LTLQAKEGLALVNGTAVSAYALRGLFYAEDLYAAA
 consensus>50 ...dd.el.a.eal...gl.p..f.LqAKEGLA\$VNGTs..ta.As.vlydanV..l.

1W27|*Petroselinum_cr* η2 α12 α13
 290 300 310 320
 1W27|*Petroselinum_cr* ...EVMSTFAEE...VMQCK..PEFTDHL..TKL.K.HHFGOIEAAIMEHTL.DG
 4C5U|*Taxus_chinensis* ...ETLQGMFCE...VIFGR..EEFAHPL..IHKV.K.PHPGOIESAELELWLL.RS
 4BAA|Redesign_1 ...ETLQGMFCE...VIFGR..EEFAHPL..IHKV.K.PHPGOIESAELELWLL.RS
 2YII|Redesign_2 ...ETLQGMFCE...VIFGR..EEFAHPL..IHKV.K.PHPGOIESAELELWLL.RS
 3NZ4|*Taxus* ...ETLQGMFCE...VIFGR..EEFAHPL..IHKV.K.PHPGOIESAELELWLL.RS
 1Y2M|*Rhododporidium_mPAL_c1* ...QSLTAXTIVE...AXVH..AGSFHPF..LHDVTR.PHPTOIEVAGNIRKLL.EG
 2NYF|*N_punctiforme* ASDDEVLSANILNFSNAENAVMQGNKSPFTDPLNI..IKLNKR.HHFGOIEAAIMEHTLDS
 2NYN|*Anabaena* ...MGVHALDIO...GLYCT..NQSFPFP..IQCK..PHPGOLWADOMFSLKDS
 1GK3|*HAL_Pseudomonas* ...MGVHALDIO...ALNGT..NQSFPFP..IENSK..PHPGOLWADOMFSLKDS
 1GKJ|*HAL_Pseudomonas* ...IACGGLSVE...AVLCSRSPPFDAR..IHEAR..GORGQIDTAACFRDLGDS
 consensus>50 ...e.l.gm.#.....vi.g...eeefahpl..iHk.k.phpgOie.Ae.me.lL.ds

ANEXO VII - Análise de qualidade dos modelos de mPAL_c1, antes e após refinamento



Avaliação da qualidade de modelo tridimensional de mPAL_c1 construído virtualmente. Avaliação da qualidade do modelo bruto por meio do *Z-Score* (A) e a análise de energia conhecida livre resíduo a resíduo (C), respectivamente. Após o refinamento da estrutura, por meio de uma metodologia híbrida, a avaliação da qualidade da mesma foi reavaliada por meio do *Z-score* (B) e análise da energia livre resíduo a resíduo (D).

ANEXO VIII - Deconvolução da curva de dicroísmo circular do agregado mPAL_c1 pelo *software* CDNN

Resumo dos principais estruturas secundárias encontradas no agregado mPAL_c1 através de dicroísmo circular. Os valores esperados (*Expected values*) foram obtidos da estrutura gerada no modelo tridimensional da Figura 19.

Estrutura	λ (nm)				<i>Expected Value</i>
	195-260	200-260	205-260	210-260	
<i>Helix</i>	12,6%	12%	9,4%	9,6%	54%
<i>Antiparallel</i>	30,7%	28,8%	29,8%	30,7%	2%
<i>Parallel</i>	5,8%	5,6%	5,5%	5,5%	1%
<i>Beta-Turn</i>	19%	19,4%	19,9%	19,2%	32%
<i>Rndm. Coil</i>	33,2%	35,2%	35,2%	35,1%	11%
Total	101,3%	101%	99,8%	100,1%	100%