

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPUS DE JABOTICABAL**

**Análise *in silico* de uma Pirofosfatase e de
uma Tioesterase de uma Biblioteca
Metagenômica de solo de *Eucalypto* spp.**

Gisele Regina Rodrigues

Bióloga

2014

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPUS DE JABOTICABAL**

**Análise *in silico* de uma Pirofosfatase e de
uma Tioesterase de uma Biblioteca
Metagenômica de solo de *Eucalypto* spp.**

Gisele Regina Rodrigues

Orientador: Prof. Dr. João Martins Pizauro Jr.

Co-orientador: Profa. Dra. Eliana G. de Macedo Lemos

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Campus de Jaboticabal, como parte das exigências para a obtenção do título de Doutora em Microbiologia Agropecuária.

2014

Rodrigues, Gisele Regina

R696a Análise *in silico* de uma Pirofosfatase e de uma Tioesterase de uma Biblioteca Metagenômica de solo de *Eucalypto* spp / Gisele Regina Rodrigues. - - Jaboticabal, 2014
xiii, 80 p. : il. ; 28 cm

Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2014

Orientador: João Martins Pizauro Junior

Banca examinadora: Alessandro de Mello Varani, Tiago Santana Balbuena, Rosa dos Prazeres Melo Furriel, Luis Henrique Souza Guimarães

Bibliografia

1. Microbiologia. 2. Biologia Molecular. 3. Metagenômica. 4. Bioinformática Estrutural. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 577.21:582.776

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.



UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE JABOTICABAL
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

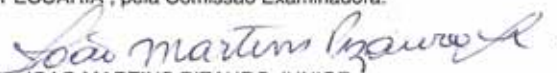
TÍTULO: ANÁLISE *in silico* DE UMA PIROFOSFATASE E DE UMA TIOESTERASE DE UMA BIBLIOTECA METAGENÔMICA DE SOLO DE *Eucalypto* spp.

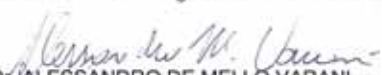
AUTORA: GISELE REGINA RODRIGUES

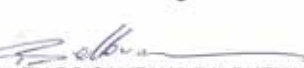
ORIENTADOR: Prof. Dr. JOAO MARTINS PIZAURO JUNIOR

CO-ORIENTADORA: Profa. Dra. ELIANA GERTRUDES DE MACEDO LEMOS

Aprovada como parte das exigências para obtenção do Título de DOUTOR EM MICROBIOLOGIA AGROPECUÁRIA, pela Comissão Examinadora:


Prof. Dr. JOAO MARTINS PIZAURO JUNIOR
Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Prof. Dr. ALESSANDRO DE MELLO VARANI
Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Prof. Dr. TIAGO SANTANA BALBUENA
Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal


Profa. Dra. ROSA DOS PRAZERES MELO FURRIEL
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto / USP / Ribeirão Preto/SP


Prof. Dr. LUIS HENRIQUE SOUZA GUIMARÃES
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto / USP / Ribeirão Preto/SP

Data da realização: 06 de maio de 2014.

DADOS CURRICULARES DO AUTOR

Gisele Regina Rodrigues – Nascida em 22 de dezembro de 1981, no município de Jaboticabal, Estado de São Paulo. Bióloga formada pelo Centro Universitário de Araraquara (UNIARA) no ano de 2005. Defendeu título de Mestre em Microbiologia na Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP- FCAV) no ano 2008.

EPÍGRAFE

"Se você quer os acertos, esteja preparado para os erros"
(Carl Yastrzemski)

AGRADECIMENTOS

Agradeço aos professores Dr. João Martins Pizauro Jr. e a Dra. Eliana Gertrudes de Macedo Lemos, por terem cedido os laboratórios L.E.A e LBMP

para realização do meu trabalho e por todo ensinamento que me proporcionaram.

Aos técnicos de laboratório Fátima Harnich e João Carlos Campanharo que me auxiliaram sempre que necessitei.

A todos meus amigos, Fernanda Andrade, Fernanda Lopes, Flavio Adriano, Rafael, Silvana, Eliamar, Camila, Andressa, Carol, que colaboraram direta ou indiretamente sempre me deram força para conseguir superar as dificuldades.

E aos meus pais Roberto e Elenice e minha irmã Kele e minhas lindas sobrinhas Marianne e Melina que me deram apoio, suporte durante toda minha vida.

Muito Obrigada!

SUMÁRIO

1. INTRODUÇÃO.....	16
2. REVISÃO BIBLIOGRÁFICA	17
2.1. Biodiversidade e Biotecnologia.....	17
2.2. Metagenômica	18
2.3 Bioinformática.....	20
2.3.2. Estrutura proteica.....	22
2.3.1. Modelagem computacional por homologia de macromoléculas	24
2.3.2. Ramachandram	27
2.4. Biocatalisadores e Biotransformação	30
2.4.2. Pirofosfatases inorgânicas (PPases)	32
2.4.3. Tioesterase (TEs).....	34
2.4.3. Família Hot- Dog.....	36
3. OBJETIVO GERAL	40
3.1. Objetivos específicos.....	40
4. MATERIAL E MÉTODOS	41
4.1. Biblioteca e sub-biblioteca metagenômica.....	41
4.1.2. Extração DNA Cosmidial.....	41
4.1.3. Reação de sequenciamento.....	42
4.1.4. Análise das Sequências.....	43
4.1.5 Seleção das ORFs	44
4.2. Identificações das proteínas hidrolíticas e análise estrutural.....	44
4.2.1 Análise Fenética	45
4.3. Amplificação e Clonagem da proteína em vetor expressão	46
4.4. Expressão e purificação de proteína	49
4.4.1. Eletroforese em gel de poliacrilamida em condição desnaturante (SDS-PAGE)	49
4.1.2. Expressão de proteínas	50
4.1.3. Purificação da proteína	50
5. RESULTADOS E DISCUSSÃO.....	51
5.1. Biblioteca, sub-biblioteca metagenômica e montagem do contig.....	51
5.2. Identificações das proteínas hidrolíticas e análise das estruturas.....	53
5.2.1. Identificações da Pirofosfatase Inorgânica (PPiase) e análise das estruturas.....	53

5.2.2. Fenograma	61
5.2.3 Identificações do gene Toiesterase (TEs) e análise das estruturas	63
5.3.1 Clonagem em vetor de expressão	73
5.3.2 Expressão e Purificação.....	74
6. CONCLUSÕES.....	76
7- REFERÊNCIASBIBLIOGRÁFICAS	77
8- APÊNDICE	86

LISTA DE FIGURAS

- Figura 1.** Comparação das abordagens metagenômicas: expressão-clonagem e clonagem em BAC para captura do material genético presente no solo onde os genes são representados por setas horizontais e os promotores pela letra P..... 19
- Figura 2.** Fluxo de análise: Esquema das principais etapas de um estudo de ecologia microbiana e metagenômica. Fonte: TEELING E GLÖCKNER, 2012..... 21
- Figura 3.** Esquema representativo dos níveis de organização estrutural das proteínas. A) representa estrutura primária, B) estrutura secundária que pode ser α -hélice ou folha- β . C) estrutura terciária, D) estrutura quaternária. 24
- Figura 4.** Grupo peptídico planar. Três ligações separam os carbonos α consecutivos em uma cadeia polipeptídica. As ligações N-C α e C α -C representam ângulos designados ϕ e ψ . Fonte: NELSON e COX, 2011... 28
- Figura 5.** Conformações proibidas entre ϕ e ψ . Algumas das restrições de rotação da ligação peptídicas. 28
- Figura 6.** Gráfico clássico do posicionamento de ϕ e ψ , tratamentos e definições. Fonte: LOVELL et al., 2003..... 29
- Figura 7.** Gráfico Ramachandran e variedade de estruturas: Os valores de ϕ e ψ de varias estruturas secundárias permitidas conhecidas cai nas regiões esperadas para as α -hélices e as conformações β . Fonte: NELSON e COX, 2011. 30
- Figura 8.** Distribuição dos tipos de enzimas utilizadas na indústria biotecnológica. As hidrolases em amarelo representam quase 50% das enzimas utilizadas em processos de biotransformação. Fonte: STRAATHOF, PANKE E SCHMID, 2002..... 31
- Figura 9.** Visualização das diferentes estruturas das Famílias PPase. A: Família I representada pelo homôdimero *S. cerevisiae*, número de acesso PDB 1E6A, B: Família II representada pelo dímero *B. subtilis*, número de acesso PDB 2HAW C: Família M-PPase representada pelo monômero *Vigna radiata* da bomba H⁺-PPase número de acesso 4A01, onde as retas horizontais em preto representam os limites intra e extracelular na membrana e a flecha azul indica o sentido do funcionamento do bomba próton do espaço intra para o extracelular . Fonte: Kanjander et al., 2013.33
- Figura 10.** Reação de catálise pela tioesterase. 34

Figura 11. Hot-dog fold: A) *E. coli* desidratase-isomerase (FabA, PDB 1MKA) a α -hélice de cor vermelha e folhas- β em verde, visto de frente. B) Giro em torno do eixo 90°. Fonte: LABONTE, TOWNSEND, 2012..... 36

Figura 12. Diferentes associações da estrutura quaternária da proteína “hot-dog (PIDUGU et al.2009)..... 37

Figura 13. Beta-oxidação da *E. coli*. O valor de 90% refere-se à percentagem de oleoil-CoA que é degradado via Beta oxidação, ao passo que 10% restante foi degradado pela via alternativa, facilitando a degradação parcial de oleoil-CoA em ácido 3,5-tetradecadienóico. Fonte: LATHAM, 2012..... 38

Figura 14. Regulação do gene *fad* e o modelo atual de sua regulação por ácidos graxos: A localização física do *fad* regulon na *E. coli*, o gene *ybaW* (*fadM*) está em cinza. B- Repressão do *fad* gene na ausência de ácidos graxos exógenos. C- Indução do *fad* gene após a adição de ácidos graxos exógenos. As estruturas ovais representam FadR, enquanto que os círculos indicam ligantes reguladores de acil-CoA e FA, ácidos graxos. Fonte: FENG e CRONAN, 2009. 39

Figura 15. Esquema ilustrativo apresentando características do vetor do sistema pET28 (NOVAGEN)..... 48

Figura 16. Organização funcional do cosmídeo: As setas indicam o tamanho relativo e a direção da transcrição de cada ORF. 52

Figura 17. Predição estrutural para MetaPPIase: onde as α -hélices estão representadas pela cor ciana, folhas- β em pink e os loops salmão. A estrutura foi criada no programa Pymol version 1.5.0.4..... 54

Figura 18. Gráfico Ramachandran baseado nos ângulos de torção ϕ e ψ : a região azul escura representa região favorável geral Pré-Pró/Prolina a azul clara região permitida geral Pré-Pró/Prolina; em laranja escuro região favorável geral para Glicina e laranja claro região permitida geral para Glicina. Gráfico foi gerado no programa Rampage..... 55

Figura 19. Pontuação de Z-score: pontos indicados com azul escuro são estruturas de proteína resolvidas por ressonância magnética nuclear (NMR) e pontos em azul claro corresponde as estruturas resolvidas através cristalografia de raios-X, o círculo preto mostra no gráfico o local da pontuação para MetaPPIase (Prosa Web)..... 56

Figura 20. Topologia transmembranar: Modelo dos segmentos transmembranares da MetaPPIase..... 57

Figura 21. Alinhamento múltiplo das sequências da H⁺PPIase: Resíduos totalmente conservados estão marcados em vermelho, e regiões com resíduos homólogos estão em marcados na caixa azul. Os resíduos Mg²⁺ e

PPi *binding* estão marcados com esferas verdes; resíduos do túnel que estão ligados no local de hidrólise e o canal de íons, indicado com esferas laranja; resíduos de abertura do canal marcado com esferas pretas e a (acc) regiões variáveis da entre as sequências representado pela cor azul. 59

Figura 22. Possível coordenada do substrato para MetaPPiase: A) Vista inclinada de um monômero Tm-PPiase do lado citoplasmático, a vista enfatiza as posições das estruturas transmembranares desordenadas, que estão numeradas de acordo com a gradiente de cores (azul para vermelho) do N para o C terminal. Fonte: Kellosalo et al 2012. B) Sobreposição das estruturas MetaPPiase em azul e Tm-PPiase laranja clara estrutura na cor laranja clara, é possível ver os ligantes de uma visão geral onde fosforo está em vermelho, Mg^{2+} laranja e K^+ roxo. B) Visão aproximada do túnel. Fonte: As figuras B e C foram desenhadas no programa Pymol version 1.5.0.4. 60

Figura 23. Mecanismo proposto para a bomba de prótons reversível de *Thermotoga marítima*: A faixa azul simboliza a membrana celular; a numeração de I a V indicam a ordem dos acontecimentos. I) Antes da ligação do substrato é possível observar o canal para a região extracelular fechado e a região intracelular aberta (sitio ativo). II) A ligação do substrato faz com que o circuito entre hélices 5 e 6 mude sua conformação, fechando a região intracelular, permitindo a formação de estado transitório onde o movimento das hélices 11 e 12 abra a região extracelular (canal de saída), e ocorre o transporte de um cátion para meio extracelular. III) O sitio ativo mantém fechado até a fim da hidrólise do pirofosfato. IV, V) Após hidrólise o grupo fosfato as duas molecular de fosfato inorgânico são levados para dentro da célula. E o canal volta para conformação inicial para uma nova hidrólise. Fonte: Kellosalo et al.,2012. 61

Figura 24. Análise Fenética MetaPPiase: As sequências de membrana PPiase foram selecionadas de acordo com o banco de dados NCBI para cada subfamília PPiase (H^+ -PPiase com K^+ independente e K^+ dependente). 62

Figura 25. Predição estrutural TEs-Meta: As estruturas em vermelho representam as α -hélices, em amarelo as folhas- β e em verde o loop. A numeração das folhas- β de 1 a 6 seguem a orientação no sentido N para o C terminal. Estrutura desenhada no programa Pymol version 1.5.0.4. 65

Figura 26. Gráfico Ramachandran baseado nos ângulos de torção ϕ e ψ para TEs- Meta: : a região azul escura representa região favorável geral Pré-Pró/Prolina a azul clara região permitida geral Pré-Pró/Prolina; em laranja escuro região favorável geral para Glicina e laranja claro região permitida geral para Glicina. Gráfico foi gerado no programa Rampage. ... 66

Figura 27. Pontuação de Z-score para TEs-Meta: pontos indicados com azul escuro são estruturas de proteína resolvidas por ressonância

magnética nuclear (NMR) e pontos em azul claro corresponde as estruturas resolvidas através cristalografia de raios-X, o círculo preto mostra no gráfico o local da pontuação para TEs-Meta. (ProSa Web)..... 67

Figura 28. Predição estrutural TEs-Meta :A) sobreposição de TEs-Meta na cor lilás e *T. thermophilus* vinho. B) Visão geral da estrutura TEs-Meta. C) Destaque para sítios de ligação, estrutura no centro representa coenzima A, resíduos ligados a CoA estão marcados com numero da posição de ligação (Thr116, Ile117, Glu118, Gly143, Ala144, Arg145 e Pro176). Estruturas desenhadas no programa Pymol version 1.5.0.4..... 68

Figura 29. Alinhamento múltiplo de sequências TEs-Meta : resíduos conservados estão marcados em vermelho, e regiões com resíduos homólogos estão em marcados na caixa em azul, o quadro vermelho marca as sequências relacionadas ao motivo da família YbgC. Acima as regiões irregulares representam em espiral as α -hélices, e as setas folhas beta. As sequências de aminoácidos usadas na análise são: 1S5U – *E. coli*, 2PZH *Helicobacter pylori*, 3DL6 *Cryptosporidium hominis*, 1YLI *Haemophilus influenzae*..... 70

Figura 30. Análise fenética de sequências proteicas hot-dog fold: As sequências foram selecionadas de acordo com similaridade genes de proteínas “hotdog fold.” As siglas D- dímeros; TA- tetrameros, TB- tetrameros com sobreposição de folhas β , H1- hexâmero com sitio ativo na interface do loop; H2 –hexâmero com hélice na interface N-Terminal; H3- hexâmero com pareamento entre dímeros e Trimêros. 72

Figura 31. Eletroforese em gel de agarose 1% dos produtos de PCR e Perfil de restrição dos plasmídeos recombinantes em vetor pET28b contendo os genes codificadores de TEs-Meta :. A: as amostras 1 e 2 são amplicons do gene TEs-Meta em duplicata e M marcador tamanho de Molecular 1 kb DNA “ladder” (Fermentas). B) 1 o é o vetor pET28a e 2 corresponde ao TEs-Meta. 73

Figura 32. Confirmação da clonagem do inserto. A sequência do gene clonado foi analisado no BlastX. 74

Figura 33. Eletroforese em SDS-PAGE a 12% dos extratos celulares de *E. coli* expressando TEs-Meta: M – marcador molecular Lonza (127 kDa to 10 kDa), as amostras 1 e 4 TEs wash 1 (lavagem) sem imidazol, amostra 2- proteína filtrada 500 mM imidazol e de 3 – 6 Proteína total e 5 – proteína filtrada e concentrada..... 75

LISTA DE TABELAS

Tabela 1. Lista dos programas mais utilizados para análise de sequências	22
Tabela 2. Programas usados para predição de estrutura e modelagem de proteínas. Fonte: SCHWEDE, 2013.....	26
Tabela 3. Famílias Tioesterase, nomes comuns dos seus membros e suas estruturas. Fonte: CANTU et al.,2010.....	35
Tabela 4. Iniciadores utilizados para fechamento do contig.	41
Tabela 5. Iniciadores para utilizados na amplificação dos genes codificadores de hidrolases.....	46
Tabela 6. Composição do gel SDS-PAGE utilizado para análise de proteínas.....	49
Tabela 7. Predição dos modelos para MetaPPIase.....	53
Tabela 8. Predição dos modelos para TEs-Meta.....	64

LISTA DE ABREVIATURAS

ABI: PRISM® 3100 GENETIC ANALYZER,
DNA: Ácido desoxirribonucleico,
D.O.: Densidade Óptica,
EAA: Arboreto *Eucalyptus* spp.,
E.C.: Comissão de Enzimas
H₊ PPIase: Bomba de Próton Pirofosfato,
H⁺ - PPIase: Íons de Hidrogênio Transportador de Pirofosfato,
IPTG: Isopropil β-D-1-tiogalactopiranosídeo,
K⁺: Íon Potássio,
kDa: Kilodalton,
LBMP: Laboratório de Bioquímica E Microrganismos de Plantas,
MetaPPIase: Pirofosfatase Inorgânica Metagenômica,
Na⁺ - PPIase: Íons de Sódio Transportador de Pirofosfato,
NCBI: Centro Nacional de Informação Biotecnologia,
ORFFINDER: Busca por Fase de Leitura Aberta,
ORFS: fase de leitura aberta,
PDB: Banco de Dados de Proteínas,
PFAM: Banco de Dados de Coleção de Famílias Proteicas,
PHRAP: Programa para Montagem de Sequências de shotgun,
PHRED: Software para Estimar Qualidade da Sequência,
PPASES: Pirofosfatase Inorgânica,
PPI: Pirofosfato Inorgânico,
ProDom: Predição Domínio de Proteínas,
PSIPRED: Análise de Sequências de Proteínas,
PYMOL: Sistema de Gráficos Moleculares,
SDS – Dodecil Sulfato de Sódio,
SDS-PAGE – Eletroforese em Gel de Poliacrilamida Desnaturante,
SP: Peptídeo Sinal,
TEs: Tioesterase,
TEs-Meta: Tioesterase Metagenômica,
TM: Transmembrana.

RESUMO: ANÁLISE *in silico* DE UMA PIROFOSFATASE E DE UMA TIOESTERASE DE UMA BIBLIOTECA METAGENÔMICA DE SOLO DE *Eucalypto* spp.

Os microrganismos apresentam uma imensa diversidade genética e desempenham funções únicas e cruciais na manutenção dos ecossistemas, uma dessas funções é a produção de enzimas extracelulares que ajudam na mineralização da matéria orgânica, liberação de carbono e nutrientes na forma de serem assimilados. Devido a esses importantes fatores cada vez mais aumenta a busca por enzimas que possam ser utilizadas nos diversos setores industriais a fim de diminuir o impacto ambiental e obter maior eficiência e menor custo durante os processos de catalise. Mas esse potencial biocatalítico ainda é pouco explorado. As hidrolases pertencem ao grupo de maior interesse para o mercado de enzimas sendo amplamente utilizadas em processos industriais tais como, a aplicação de aditivos em alimentos para animais, têxteis, para a produção de biodiesel e biossíntese de novos antibióticos. Nesse sentido, o objetivo do trabalho foi o de encontrar novas enzimas hidrolíticas. Para isso, foi utilizada biblioteca metagenômica de solo sob *Eucalyptus* spp. (EAA) realizando uma busca para identificar clone positivo e a construção da biblioteca “shotgun”. Os dados gerados pelo sequenciamento foram utilizados para a montagem do mapa com sequências conservadas que permitiu identificação de genes codificadores de pirofosfatase inorgânica (PPase) e tioesterase (TEs). A pirofosfatase inorgânica tem um papel importante no metabolismo do fósforo, além de controlar a concentração celular de pirofosfato (PPi) e processo de biossíntese. Tioesterases (TEs) são responsáveis pela hidrólise de um grupo tioéster carboxílico e um átomo de enxofre em ácidos graxos. A metagenômica revelou-se como uma ferramenta inovadora para explorar novos biocatalisadores na microbiota presente no solo. A análise dos dados *in silico* permitiu a predição da estrutura tridimensional e possível função das proteínas, assim os resultados sugerem que MetaPPase é uma fosfatase integral de membrana do grupo H + _PPase e subfamília de K⁺ independente e a TEs-Meta foi classificada como membro tioesterase tipo II pertencente ao grupo Hotdog-fold associada a subfamília YbgC .

Palavras-Chave: diversidade microbiana do solo; bioinformática e expressão de proteína.

ABSTRACT: *IN SILICO* ANALYSIS OF A PYROPHOSPHATASE AND A THIOESTERASE FROM A SOIL METAGENOMICS LIBRARY *EUCALYPTUS* SPP.

The soil diversity of microorganisms with potential to select new genes, enzymes and compounds able to supply the high demand of biotechnology industry. In this sense, an immense diversity of microorganisms, yet most remains unexplored. Thus, soil diversity it is stimulated the search for new biocatalyst to high efficiency and stability, has been a constant pursuit of enzymologists, biochemists and chemical engineers. One of that biocatalyst is the hydrolases group, which are great interest to market for enzymes that cleavage of chemical structure by the addition of water. The hydrolases are widely used in industrial processes such as the application of additives in animal feed, textile, biodiesel production and biosynthesis of new antibiotics. Thus, the aim of this work was to find new hydrolases genes. In this paper, we used metagenomic library of soil under *Eucalyptus* spp. arboretum (EAA), was performed screening to identify a positive clone and construct the shotgun library. The data was analyzed using bioinformatics tools, to assemble the functional map and identify a hydrolase gene as possible novel inorganic (PPase) a thioesterase (TEs). The inorganic pyrophosphatase that has an important role in phosphorus metabolism controlling the cellular concentration of pyrophosphate (PPi) and biosynthetic process. Thioesterase (TEs) responsible for the hydrolysis of a carboxylic thioester group and a sulfur atom in fatty acids. The metagenomic revealed itself as an innovative tool for exploring novel biocatalysts in the microbiota present in the soil. The data analyses provide three-dimensional structure and protein function and the results suggest that MetaPPase is M- PPase and perchance an H⁺_PPase transporting subfamilies and TEs_ Meta possibly a new thioesterase.

Keywords: soil microbial diversity; bioinformatics and protein expression.

1. INTRODUÇÃO

Os microrganismos encontram-se amplamente distribuídos na biosfera e apresentam imensa diversidade genética e desempenham funções essenciais na manutenção dos ecossistemas, como elementos fundamentais de cadeias alimentares e ciclos biogeoquímicos (WHITMAN et al., 1998).

Estudos indicam que o solo possui uma grande diversidade microbiana, porque apenas uma pequena parcela de microrganismos do solo é isolada pelas técnicas tradicionais de cultivo. Entretanto, ferramentas moleculares e tecnologias baseadas em sequências gênicas vêm reduzindo essas limitações e revelando nova perspectiva sobre a diversidade microbiana (PETTIT, 2004; RONDON et al., 2000; STREIT e SCHMITZ, 2004). A metagenômica é a estratégia que permite acessar mais informação genética que os procedimentos baseados técnicos tradicionais de cultivo (HANDELSMAN, 1998). A extração DNA genômico é feita a partir de uma amostra ambiental seguido por a construção de bibliotecas e triagem dos genes de interesse permitindo a identificação e isolamento de genes completos que possam sintetizar novas enzimas independente de cultivo (IQBAL et al., 2012; LEE et al., 2013; SIMON e DANIEL, 2009). Nesse sentido, o conhecimento da diversidade microbiana do solo aumenta a busca de novos biocatalisadores com alta eficiência e estabilidade durante processos industriais. Onde os biotransformantes mais utilizados são: oxidorreduções, liases, isomerases, redutases e hidrolases (STRAATHOF et al., 2002). Sendo as hidrolases o grupo de enzimas mais pesquisado para a atuação como biocatalisadores nos diversos setores da indústria como: cosmético, ração, bebidas, têxtil, produção de biodiesel e biossíntese de novos antibióticos (LEE et al., 2013; RODRÍGUEZ-GUILBE et al., 2013).

Dentre as hidrolases, as tioéster hidrolase ou tioesterase (TEs EC 3.1.2.1), e a pirofosfatase inorgânica (PPIase EC 3.6.1.1) são de grande interesse para o mercado de enzimas, por serem enzimas ubíquas e participarem reações metabólicas importantes.

Tioesterase (TEs EC 3.1.2.1) são enzimas responsáveis pela hidrólise das ligações de tioéster entre um grupo carboxílico e um átomo de enxofre em ácidos graxos (DILLON & BATEMAN, 2004; PIDUGU et al., 2009; RODRÍGUEZ-GUILBE, et al., 2013). A pirofosfatase inorgânica (PPiase EC 3.6.1.1) é a enzima central do metabolismo do fósforo sendo essencial, para hidrólise do pirofosfato inorgânico (PPi) gerando fosforo inorgânico (Pi) e ATP que fornece energia termodinâmica para os processos celulares (FABRICHNIY et al., 2006; KELLOSALO et al., 2012; KAJANDER et al., 2013).

Diante do potencial biocatalítico dos microrganismos do solo e frente às limitações de cultivos, esse estudo utilizou abordagem metagenômica para identificar novas hidrolases com maior especificidade. Com este propósito, uma biblioteca metagenômica de solo sob arboreto de *Eucalypto spp.* foi utilizada para selecionar novas proteínas com atividade hidrolítica as quais foram submetidos a análise *in silico* para prospecção estrutural e funcional das proteínas e subsequente expressão em sistemas heterólogos.

2. REVISÃO BIBLIOGRÁFICA

2.1. Biodiversidade e Biotecnologia

Nos últimos 50 anos, produtos derivados do metabolismo secundário microbiano têm sido utilizados para fins médicos, industriais e agrícolas (OLIVEIRA et al.; 2006). A maioria dos metabólitos, em uso, é obtida a partir de microrganismos habitantes do solo (HUNTER-CEVERA, 1998; RAAIJMAKERS et al., 2002).

Entretanto, a diversidade é tão grande quanto desconhecida a principal razão para isso, é que até pouco tempo os microrganismos precisavam ser isolados e classificados através de técnicas tradicionais de cultivo (PACE, 1997). Mas alguns microrganismos possuem condições especiais para seu desenvolvimento como à sazonalidade, e em muitos casos, a dependência de hospedeiros e/ou substratos específicos para sua

sobrevivência e multiplicação, atuando como fatores limitantes para identificação de microrganismos não isolados através das técnicas tradicionais (TORSVIK and ØVREÅS, 2002).

2.2. Metagenômica

A metagenômica permite examinar a composição funcional e a diversidade de comunidades microbianas em seus habitats naturais, e permite o acesso ao conteúdo genômico da maioria dos organismos não cultiváveis (HANDELSMAN et al., 1998; HANDELSMAN, 2004), de um determinado ambiente através do sequenciamento do DNA e/ou RNA (PRAKASH e TAYLOR, 2012), oferecendo um caminho para acessar mais completamente a diversidade microbiana total de um solo, permitindo examinar a genômica funcional dos membros de sua microbiota, mesmo na ausência de cultivo de muitos dos organismos presentes.

Assim para acessar potencial metabólico é necessária a construção de bibliotecas metagenômicas, onde fragmentos de DNA, extraídos de uma amostra ambiental, são clonados em vetores do tipo cosmídeos, fosmídeos ou BACs, vetores estes, que suportam a clonagem de fragmentos grandes de DNA (20-300 Kb).

Uma vez que estes DNAs estejam imobilizados em vetores são analisados em busca de novos produtos com atividades biológicas desejadas (HANDELSMAN et al., 1998; RONDON et al., 2000) oferecendo dessa forma uma alternativa de acesso mais abrangente a diversidade microbiana de um dado ambiente, permitindo a análise de genes (Figura 1).

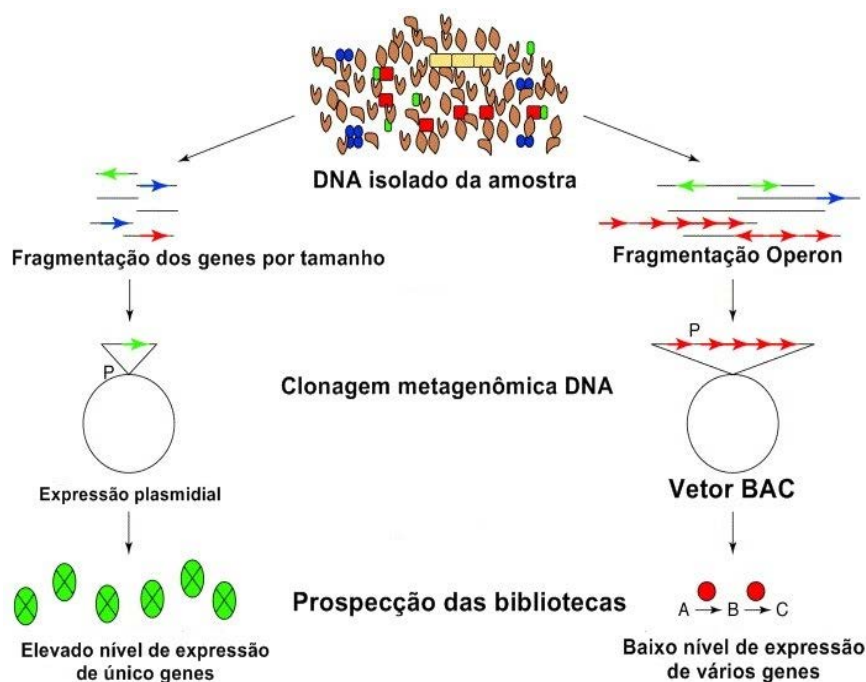


Figura 1. Comparação das abordagens metagenômicas: expressão-clonagem e clonagem em BAC para captura do material genético presente no solo onde os genes são representados por setas horizontais e os promotores pela letra P.

A caracterização completa de um fragmento de DNA clonado implica na determinação da sua sequência de nucleotídeos. Um dos métodos utilizados para este fim é o método de Sanger, o qual permite determinar a sequência exata de uma cadeia de DNA com leitura de até 1000 nucleotídeos.

O método consiste na síntese enzimática de uma cadeia complementar de DNA usando um iniciador e didesoxinucleotídeos (ddNTPs) marcados com diferentes marcadores fluorescentes (NELSON e COX, 2011).

Nesse sentido, a metagenômica facilitou a busca por novos genes de interesse biotecnológico, além de tornar possível a exploração dos mais diversos ambientes como: regiões vulcânicas continentais (STOUT et al., 2009), profundidades oceânicas (DICK e TEBO, 2010; SCHECKENBACH et al., 2010), geleiras na Antártida (CARY et al., 2010, PEARCE et al., 2010, TEIXEIRA et al., 2010), fontes termais continentais distribuídas por todo o mundo (SAHA et al., 2010, SONG et al., 2010, VICK et al., 2010), e mais recentemente, a microbiota humana, em particular, do trato

gastrointestinal ganhou considerável atenção, além do interesse em caracterizar a microbiota de diferentes locais do corpo, com objetivo final de compreensão da saúde e das doenças humana (MENDE et al., 2012) entre outras.

2.3 Bioinformática

Embora existam diferentes metodologias para o sequenciamento de bibliotecas metagenômicas, a análise dos dados gerados a partir do sequenciamento possui uma ordem de passos independente da metodologia. Com o aumento de genomas e metagenôma sendo sequenciados houve aumento considerável na quantidade de sequências armazenadas em banco de dados, demandando um aperfeiçoamento constante de recursos computacionais.

Nesse sentido, com a utilização de ferramentas de bioinformática é possível interpretar grande volume de dados gerados a partir de pesquisas na área de genoma (PRAKASH e TAYLOR, 2012).

A etapa de montagem do genoma ou metagenôma consiste na geração da sequência genômica a partir dos fragmentos já sequenciados. Os programas utilizados são baseados em métodos heurísticos e de programação dinâmica de alinhamento de sequências que essencialmente procuram encontrar redundâncias entre fragmentos, agrupando os que tenham uma parte comum, que formam fragmentos maiores, chamados fragmentos consensos ou contigs (MOUNT, 2004).

Existem diferentes maneiras para certificar que uma sequência genômica esta correta (ALBERTS et al., 2002). A sequência pode ser comparada com pequenas partes de um genoma sequenciado e anotado previamente ou com regiões de mapas genômico. Se uma montagem é consistente com regiões dispersas de informação conhecida, então é bastante provável que toda a sequência possa estar correta (STEIN, 2001). A etapa de anotação de genoma é crítica porque consiste na identificação de regiões onde estão localizados os possíveis genes para depois determinar quais são suas funções e descrições biológicas (STEIN, 2001) a

partir desta etapa as análises e ferramentas são específicas para o objetivo da biblioteca em análise que pode ser: prospecção de genes, análise funcional e metabólica, análise comparativa e classificação taxonômica das sequências (Figura 2).

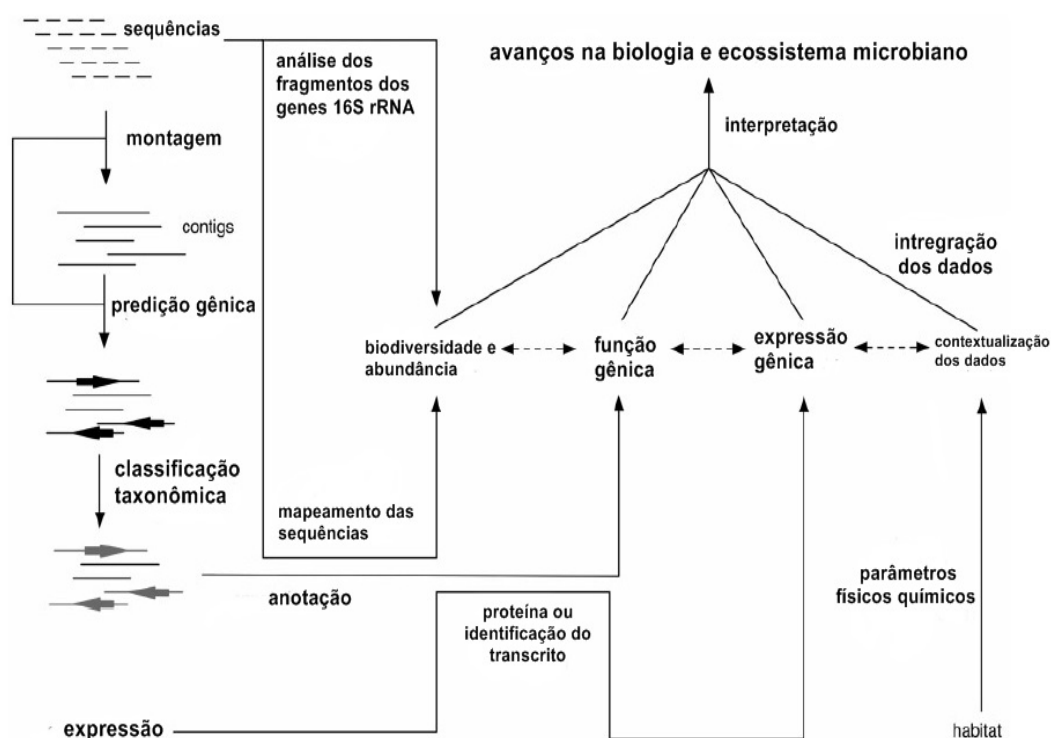


Figura 2. Fluxo de análise: Esquema das principais etapas de um estudo de ecologia microbiana e metagenômica. Fonte: TEELING E GLÖCKNER, 2012.

A figura 3 ilustra as diferentes análises que podem ser feitas ao montar uma biblioteca ou genoma e a sequência de passos devem se seguidos, com os dados obtidos após o sequenciamento. A tabela 1 mostra alguns dos programas utilizados para as diferentes abordagens metagenômicas utilizados atualmente.

Tabela 1. Lista dos programas mais utilizados para análise de sequências

FUNÇÃO	PROGRAMA	URL/ Referência
Montagem de Sequências	Phrap	http://www.phrap.org
	Forge	http://www.combiol.org/froge/
	Arachne	BATZOGLOU et al.,2002
	JAZZ	APARICIO et al.,2002
	Celera	MAYERS et al.,2000
	Velvet	ZERBIO et al.,2008
	Newbler	454 Life Sciences
	SOAPdenovo	LI et al.,2010
	EULER	PEVZNER et al.,2001
	ORFome assembly	YE AND TANG 2009
	IDBA-UD	PENG et al.,2012
Predição Gênica	Metagene	NOGUCHI et al.,2006
	GeneMark	ZHU et al., 2010
	Orf finder	http://www.ncbi.nlm.nih.gov/projects/gorf
	Frag Gene Scan	RHO et al., 2010
	fgenesB	http://www.softberry.com
	GLIMMER	DELCHER et al.,1999
	BLAST	ALTSCHUL et al., 1990
Predição de genes de RNA	tRNAscan-SE	LOWE et al.1997
	Similaridade baseada na busca em banco de dados rRNA	-
Análise Taxonômica	MetaBim	SHARMA et al., 2012
	MEGAN	HUSON et al., 2007
	WebCARMA	GERLACH et al.,2009
	PhyloPythia	MCHARDY et al., 2007
	TETRA	TEELING et al., 2004
	NBC	ROSEN et al.,2011
	TACOA	DIAZ et al., 2009

2.3.2. Estrutura proteica

As propriedades funcionais biológicas das proteínas dependem de seu dobramento correto em estruturas secundárias, terciárias e quaternárias (quando presentes), que lhes conferem uma conformação funcional. A conformação estrutural definirá estrutura e função das

proteínas que podem ser: fibrosas, globulares, transportadores de membrana celular, estrutural, hormônio, anticorpo ou enzima (ALPHEY, 1997).

Para compreender a estrutura proteica é necessário descrever e compreender as estruturas que compõem os diferentes níveis de organização estrutural. A estrutura primária (Figura 3A) é o primeiro determinante das interações específicas que permitem que um pequeno conjunto de pouco mais de 20 aminoácidos dê origem a uma quantidade imensurável de proteínas nos seres vivos. Isso porque as características distintas das cadeias laterais destes aminoácidos que possuem diferenciações quanto a presença de carga elétrica (carregados positiva ou negativamente), potencial hidrofílico / hidrofóbico, presença de grupos aromáticos. Como os mesmos se encontram ligados por ligações peptídicas, a proximidade destas cadeias laterais irá fazer com que esses resíduos interajam entre si, repelindo ou atraindo-se mutuamente por interações eletrostáticas, hidrofóbicas, hidrofílicas, pontes de dissulfeto, e atuação de forças de “van der Waals”. A estrutura secundária é determinada pelo arranjo local dos aminoácidos em cadeia – alfa-hélices e folhas- β (60% das estruturas mais frequentes), *loops*, *turns*, etc. (Figura 3B). As estruturas secundárias por sua vez, também interagem entre as interações eletrostáticas, hidrofóbicas, hidrofílicas, pontes de dissulfeto, e atuação de forças de “van der Waals”, auxiliam na formação de dobramentos mais complexos como a estrutura terciária. (Figura 3C). A interação entre duas ou mais estruturas terciárias dará origem a estrutura proteica quaternária, como pode ser observado na (Figura 3D).

Estes empacotamentos das estruturas, auxiliados por modificações pós-traducionais (pós-sintéticas, como a adição de grupos químicos ligantes funcionais e a substituição de aminoácidos comuns por não usuais) irão garantir o papel biológico da mesma nos organismos vivos. (NELSON E COX, 2011).

ferramenta mais eficiente na predição de estruturas tridimensionais de proteínas é a modelagem por homologia, também conhecida como modelagem comparativa (*comparative protein modeling*) (SIMONS et al.,1997, HANS-DIETER, H., & FOLKERS,2008). Esta abordagem baseia-se na homologia entre sequências de aminoácidos implicando em semelhança estrutural e funcional; proteínas homólogas apresentam regiões internas conservadas (principalmente constituídas de elementos de estrutura secundária: α -hélices e folhas- β ; as principais diferenças estruturais entre proteínas homólogas ocorrem nas regiões externas, constituídas principalmente por alças (*loops*), que ligam os elementos de estruturas secundárias. (SANTOS FILHO e BICCA DE ALENCASTRO, 2003).

Apesar da complexidade das sequências das proteínas a modelagem computacional é possível porque proteínas homologas compartilham similaridades entre as sequências (domínios conservados), além de apresentar a estrutura 3D semelhante (SCHWEDE, 2013).

A conservação de resíduos em proteínas homólogas é notável, pois sequências de resíduos de aminoácidos em torno de 30% de identidade podem ter excelente sobreposição (*protein backbone*), com desvios de mínimos quadrados (rmsd) da ordem de 2 Å, comparáveis aos valores de “rmsd” da ordem de 0,7 Å encontrados em proteínas idênticas em diferentes formas cristalinas (BENNER et al.,1997, SCHWEDE, 2013). Em casos em que a identidade entre as sequências é inferior a 25%, a melhor maneira de resolver esse problema é utilizar métodos de compatibilidade sequência-estrutura (*threading* ou *3D template matching methods*) (BOWIE et al., 1991)

Atualmente existem diferentes programas que permitem realizar a modelagem de proteínas por homologia, alguns desses programas estão indicados na tabela 2 (GUEx et al., 2009).

Tabela 2. Programas usados para predição de estrutura e modelagem de proteínas. Fonte: SCHWEDE, 2013.

Programas	Web site
Protein Model Portal	http://www.proteinmodelportal.org (Arnold et al., 2009; Haas et al., 2013)
Model Archive	http://modelarchive.org
HHpred	http://toolkit.tuebingen.mpg.de/hhpred (Hildebrand et al., 2009)
IMP	http://www.salilab.org/imp (Russel et al., 2012; Yang et al., 2012)
IntFOLD	http://www.reading.ac.uk/bioinf/IntFOLD/ (Roche et al., 2011)
I-Tasser	http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Zhang, 2013)
ModBase	http://salilab.org/modbase/ (Pieper et al., 2011)
Modeler/ModWeb	http://salilab.org/modeller/ (Pieper et al., 2011; Yang et al., 2012)
Pcons.net	http://pcons.net/ (Larsson et al., 2011)
PHYRE2	http://www.sbg.bio.ic.ac.uk/phyre2/ (Kelley and Sternberg, 2009)
Robetta	http://rosetta.bakerlab.org/ (Raman et al., 2009)
Rosetta	https://www.rosettacommons.org (Das and Baker, 2008)
SWISS-MODEL Repository	http://swissmodel.expasy.org/repository (Kiefer et al., 2009)
SWISS-MODEL Workspace	http://swissmodel.expasy.org/workspace/ (Arnold et al., 2006; Bordoli and Schwede, 2012)

É importante ressaltar que após a modelagem computacional devem ser realizados testes de confiabilidade em relação às estruturas reais. Para isso, é necessário checar resíduos conservados para o sítio ativo proteico, verificando a correspondência dos padrões destas regiões entre o modelo e o molde.

Adicionalmente, os padrões de características espaciais do modelo compilados de estruturas conhecidas de alta resolução podem ser testados por programas que trabalham com métodos baseados no perfil 3D e potenciais estatísticos de força. Os valores de R.M.S.D. (programa CCP4 e VERIFY3D) auxiliam na identificação e confiança do modelo por meio da identificação do Score S_{ideal} . Valores de S_{ideal} abaixo de 0,45 indicam modelos com erros de predição no enovelamento, e valores próximos ou acima destes indicam estruturas confiáveis. Analisados em conjunto, estes valores irão indicar quais as informações biológicas que podem ser inferidas a partir de um dado modelo, garantidas que as extrapolações

feitas a partir dos resultados assumam proporções confiáveis, limitados à capacidade preditiva do modelo.

Nesse sentido, a modelagem 3D *in silico* de proteínas é um método alternativo que vem auxiliar os pesquisadores na predição de estruturas 3D de proteínas e reconhecimento de sítios ativos. Segundo Schwede (2013), a ampla disponibilidade de informações de estrutura 3D permite abordagens racionais baseadas na estrutura para as mais diversas aplicações nas pesquisas.

2.3.2. Ramachandram

Utilizando o gráfico de Ramachandran é possível avaliar a qualidade estereoquímica do modelo, refletida pelos padrões adequados de comprimento de ligação, ligação peptídica e a estrutura planar de anéis de cadeia lateral, quiralidade, ângulo de torção para a cadeia principal e cadeia lateral, além de choques entre pares de átomos não ligados (MORRIS et al., 1992, LOVELL et al., 2003), que permite visualizar a disposição espacial do modelo em regiões mais favoráveis, com base na conformação molecular com menor energia livre no sistema.

As proteínas são formadas por ligações peptídicas que ocorrem entre o grupo α -carboxila de um aminoácido e o grupo α -amino de outro aminoácido. A cadeia polipeptídica é formada por uma sucessão de planos peptídicos, onde o único ponto de rotação possível da cadeia peptídica é em torno do $C\alpha$. E os ângulos de torção entre ϕ (fi) (ângulo de rotação entre $C\alpha$ e o N) e ψ (psi) (ângulo de rotação entre $C\alpha$ e o C), descreve a orientação relativa dos planos peptídicos (Figura 4) (NELSON e COX, 2011).

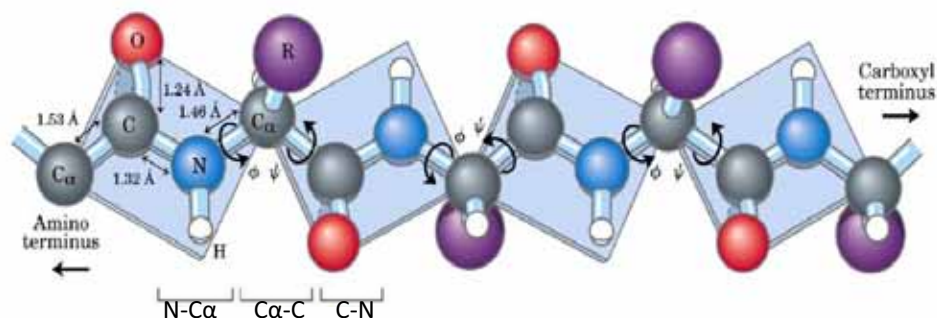


Figura 4. Grupo peptídico planar. Três ligações separam os carbonos α consecutivos em uma cadeia polipeptídica. As ligações N-C α e C α -C representam ângulos designados ϕ e ψ . Fonte: NELSON e COX, 2011.

A princípio os ângulos ϕ e ψ podem ter qualquer valor entre $+180^\circ$ e -180° (Figura 5). Onde as rotações das ligações peptídicas implicam em colisão da nuvem eletrônica do O ou N (Figura 5 B e D), ou ainda no posicionamento dos átomos de tal modo que não possibilita a formação de pontes de H (Figura 5 A e C).

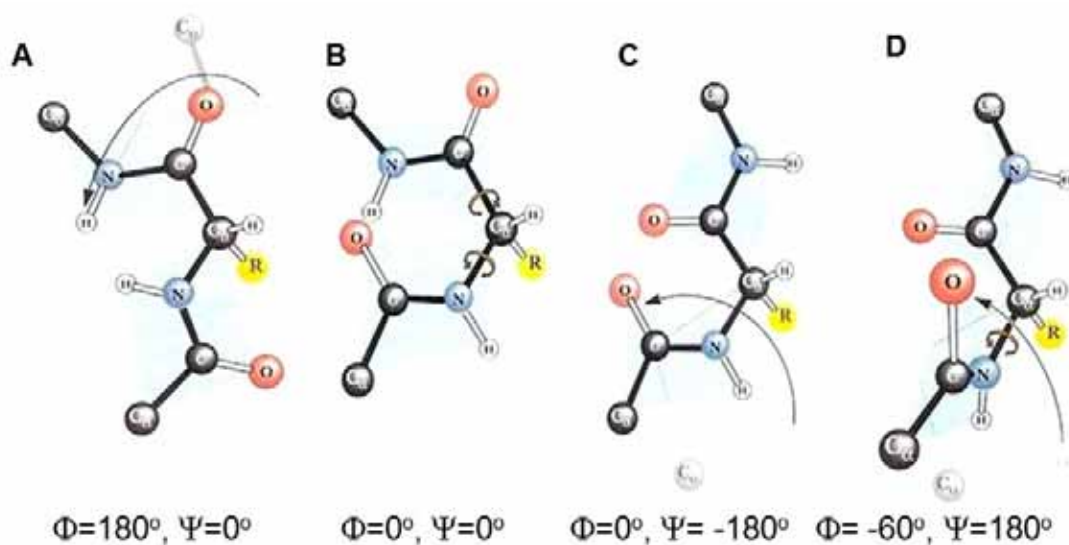


Figura 5. Conformações proibidas entre ϕ e ψ . Algumas das restrições de rotação da ligação peptídica.

No entanto, diversos valores são proibidos por impedimento estérico entre os átomos da cadeia peptídica e os das cadeias laterais dos aminoácidos, sendo que, os valores permitidos por ϕ e ψ tornam-se evidentes quando ψ é colocado em um gráfico versus ϕ no diagrama de

Ramachandran (Figura 6 A). Lovell et al (2003) propôs melhoras para o gráfico de Ramachandran onde as regiões permitidas para dos átomos foram ampliadas, tornando a validação estrutural mais eficaz (Figura 6B).

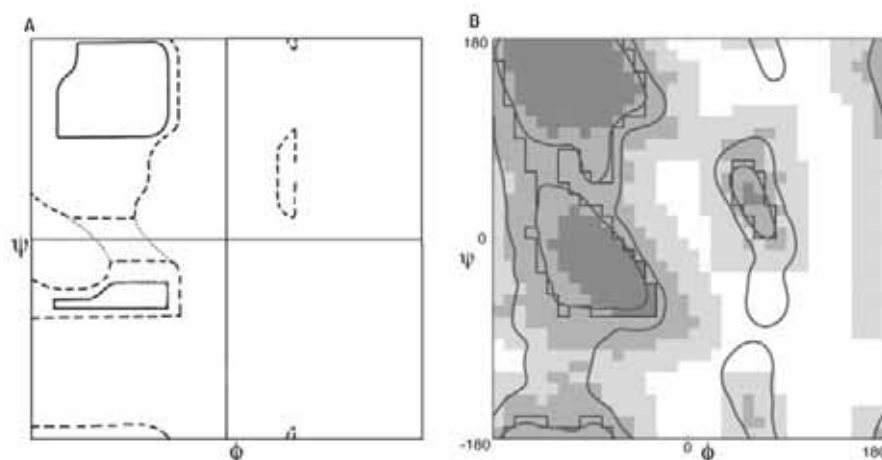


Figura 6. Gráfico clássico do posicionamento de ϕ e ψ , tratamentos e definições. Fonte: LOVELL et al., 2003.

Na figura 6 A, a região delimitada por linhas continua é região permitida sem sobreposição atômica, linhas tracejadas representam regiões permitidas indo ao limite extremo da sobreposição, linhas pontilhadas regiões permitidas com menor flexibilidade entre os ângulos. E figura 6 B, Região que destacada por cinza escura representa região permitida se sobreposição de resíduos, cinza médio limite extremo da sobreposição resíduo com raios ligeiramente menores e cinza claro regiões permitidas com menor flexibilidade entre os ângulos

A função das proteínas está relacionada com a sua conformação tridimensional e esta conformação, por sua vez, é determinada pela sequência de aminoácidos, os quais possibilitam a formação de diferentes tipos de estruturas secundárias, e estão associados aos diferentes ângulos ϕ e ψ , a cada tipo corresponde uma região específica do diagrama de Ramachandran (LOVELL et al., 2003) (Figura7).

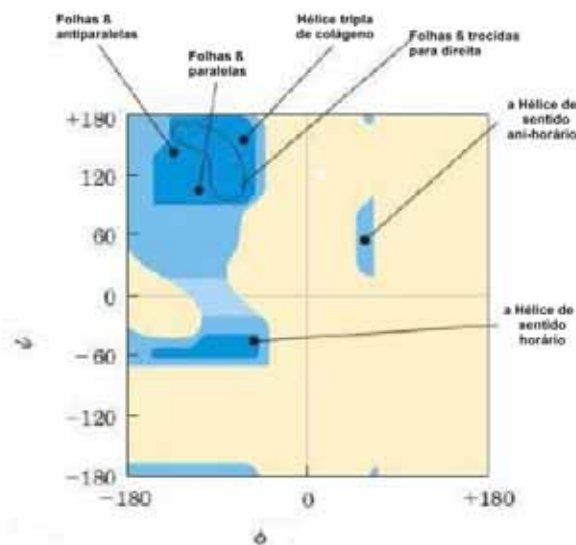


Figura 7. Gráfico Ramachandran e variedade de estruturas: Os valores de ϕ e ψ de varias estruturas secundárias permitidas conhecidas cai nas regiões esperadas para as α -hélices e as conformações β . Fonte: NELSON e COX, 2011.

Além disso, existem dois aminoácidos com características especiais em relação ao gráfico de Ramachandran: **Prolina** – o seu grupo amina não é livre, encontrando-se ligado ao radical, o que confere grande rigidez induzindo quebra nas α -hélices e nas folhas- β ; **Glicina** – pode ser encontrada em qualquer lugar do diagrama de Ramachandran, nas proteínas é muitas vezes encontrada em regiões flexíveis.

2.4. Biocatalisadores e Biotransformação

Os metabolitos secundários produzidos pelos microrganismos são responsáveis por uma série de transformações químicas com as quais o ser humano tem tomado contato desde sua primitiva existência.

As enzimas sintetizadas pelos microrganismos podem promover reações em condições brandas de pH, temperatura e pressão, características essas que fazem das enzimas catalizadores altamente específicos (CORREA & ZUIN, 2012).

O termo biocatálise aplica-se as reações catalisadas por enzimas isoladas ou contidas em organismos biológicos, estudos revelaram que

processos catalisados por enzimas são mais rápidos e com maior rendimento, quando comparados aos químicos (SANTOS et al., 2012).

A produção industrial de substâncias como fármacos, agroquímicos, complementos alimentares, perfumes dentre tantos outros bens de consumo implica na poluição e contaminação ambiental. Nesse sentido as indústrias estão substituindo gradualmente o uso de reagentes químicos por enzimas visando diminuir o impacto ambiental e obter maior eficiência e menor custo (SARROUH et al., 2012).

O mercado mundial de enzimas industriais movimentou um valor estimado de 3,3 bilhões de dólares em 2010 e deverá chegar a 8,0 bilhões até 2015. As classes de enzimas utilizadas nos diversos setores da indústria são: hidrolases, oxidorreduções, transferases, liases, ligases e isomerases. (SARROUH et al., 2012; BINOD et al., 2013) e as hidrolases respondem por quase metade do mercado de enzimas (Figura 8).

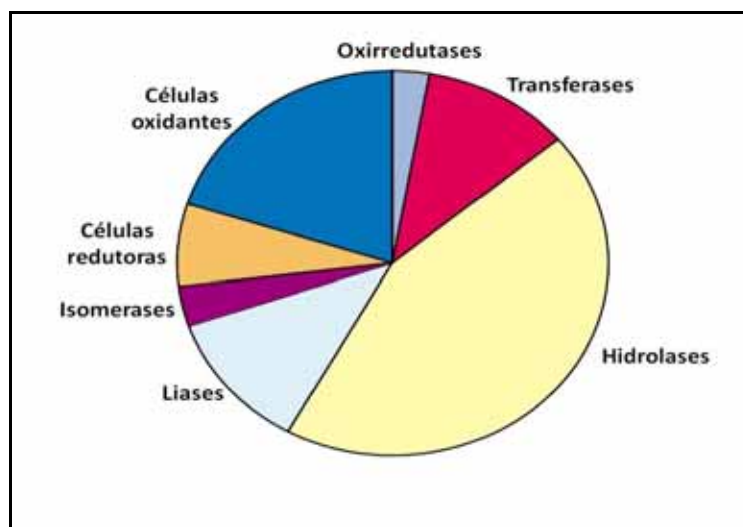


Figura 8. Distribuição dos tipos de enzimas utilizadas na indústria biotecnológica. As hidrolases em amarelo representam quase 50% das enzimas utilizadas em processos de biotransformação. Fonte: STRAATHOF, PANKE E SCHMID, 2002.

As hidrolases são enzimas que hidrolisam ligações C-O C-N e C-C, entre outras e são representadas por: lipases, protease, fosfatases, amilases. Assim as hidrolases são catalisadores biológicos capazes de interagir com uma grande variedade de substratos. A sua aplicação no

processamento de materiais biológicos apresenta muitas vantagens em relação à hidrólise química em condições de reação mais brandas e com menor gasto energético (FERRER, MARTÍNEZ-ABARCA e GOLYSHIN, 2005).

2.4.2. Pirofosfatases inorgânicas (PPases)

As Pirofosfatases inorgânicas (PPase, E.C 3.6.1.1) encontram-se amplamente distribuídas em plantas, bactérias, protistas e archaeas, sendo considerada enzima central do metabolismo celular, atuando no controle da concentração intra e extracelular do pirofosfato inorgânico (PPi), além de controlar as reações biosintéticas como por exemplo: ácido nucléico e proteínas (KAJANDER et al., 2013; BAYKOV et al., 2013; LUOTO et al., 2013). As PPases também são responsáveis por hidrolisar o pirofosfato inorgânico (PPi) durante o transporte ativo de cátions através da membrana utilizando o PPi como fonte de energia, em vez de ATP (KAJANDER et al., 2013; LUOTO et al., 2013; WANG et al., 2013; FABRICHNY et al., 2007). As PPases são classificadas de acordo com a função biológica e pH, divididas em famílias I e II que representam as solúveis e de membrana (M-PPase) (KAJANDER et al., 2013; BAYKOV et al., 2013; JAMSEN et al., 2012).

A família I é comumente encontrada em todos os tipos de organismos e a família II tem sido encontrada até o momento em bactérias do gênero *Bacillus* e *Clostridia* além de alguns patógenos humanos (YOUNG et al., 1998; SHINTANI et al., 1998) e algumas linhagens de archaea. As PPases solúveis mais estudadas são a hexamérica de *Escherichia coli* e a dimérica de *Saccharomyces cerevisiae*. Estas enzimas têm sido extensivamente caracterizadas por cristalografia de raios-X, mutações sítio-dirigido, bem como o mecanismo de hidrólise (FABRICHNIY et al., 2007; OKSANEN et al., 2007; KAJANDER et al., 2013).

Os membros da família I estudadas até o momento requerem cátions metálicos como cofatores. Contudo, as PPases de bactérias, levedura e mamíferos foram descritas por serem fortemente inibidas por Ca^{2+} a concentrações fisiológicas (FELIX E FLEISCH, 1975; KURILOVA et al.,

1984; MITCHELL e MINNICK, 1997). Seu mecanismo de inibição foi investigado por SAMYGINA et al.,(2001) onde se concluiu que o Ca^{2+} impede o ataque nucleofílico de água que participa da rede de coordenação ao pirofosfato (Figura 9 A)

Funcionalmente, a família II difere da família I pela sua preferência por Mn^{2+} ao invés de Mg^{2+} como cofator. O Mg^{2+} é o melhor cofator para as PPases da família I e ele também atua na mesma proporção as PPases da família II, contudo, o Mn^{2+} confere uma atividade 20 vezes maior para as PPases da família II do que o Mg^{2+} (Figura 9B) (PARFENYEV et al., 2001; ZYRYANOV et al., 2004).

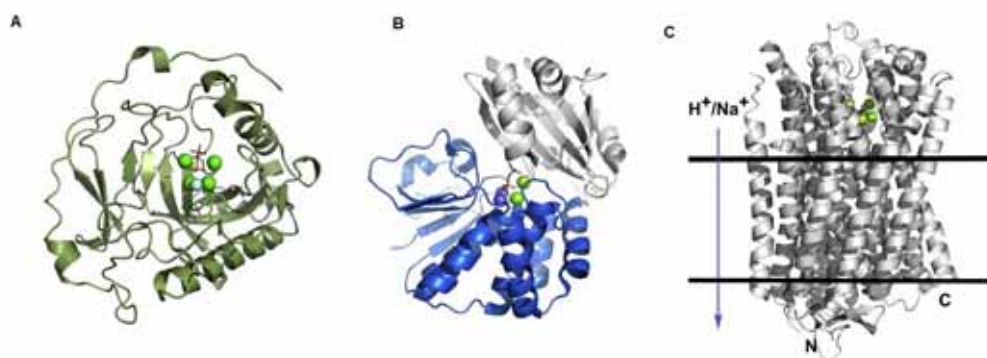


Figura 9. Visualização das diferentes estruturas das Famílias PPase. A: Família I representada pelo homôdimero *S. cerevisiae*, número de acesso PDB 1E6A, B: Família II representada pelo dímero *B. subtilis*, número de acesso PDB 2HAW C: Família M-PPase representada pelo monômero *Vigna radiata* da bomba H^+ -PPase número de acesso 4A01, onde as retas horizontais em preto representam os limites intra e extracelular na membrana e a flecha azul indica o sentido do funcionamento do bomba próton do espaço intra para o extracelular. Fonte: Kanjander et al., 2013.

As pirofosfatases de membrana (M-PPiase) possuem uma arquitetura e função completamente diferente das famílias já descritas (Figura 9). As M-PPiases atuam como uma bomba de prótons reversível. E através da membrana celular produzindo gradiente de próton e/ou sódio, em plantas, algas e em alguns protistas, bactérias e archaeas. Sua estrutura possui grande domínio transmembranar com 15 a 17 hélices transmembrânicas (TM), com grande região hidrofílica do lado citoplasmático da proteína. Todas as M-PPiases até agora estudadas são homodiméricas, embora o sítio ativo fique contido dentro de um único monômero (KAJANDER et al.,2013, KELLOSALO et al., 2012). A estrutura da M-PPiase foi

desvendada recentemente por LIN et al., 2012 e KELLOSALO et al., 2012 que utilizaram respectivamente como modelos para seus estudos PPIases de feijão verde *Vigna radiata* e uma bactéria termófila *Thermogota maritima*, ainda existem varias lacunas a serem preenchidas em relação a estrutura e funcionamento da bomba de próton reversível.

2.4.3. Tioesterase (TEs)

Tioéster hidrolase ou tioesterase (TEs EC 3.1.2.1) desempenham papéis celulares importantes na produção de energia, ciclo celular, transdução de sinal e regulação gênica, além de catalisar a hidrólise do acil-CoA tioéster em ácidos graxos livres e coenzima A (CoA-SH) (Figura 10). (DILLON e BATEMAN, 2004; PIDUGU et al. 2009; RODRÍGUEZ-GUILBE, et al. 2013).

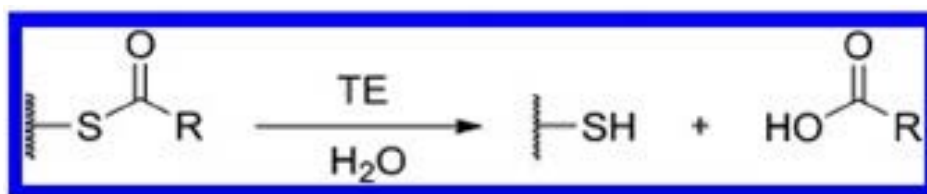


Figura 10. Reação de catálise pela tioesterase.

Existem duas famílias de enzimas que evoluíram para catalisar a hidrólise dos tioésteres, são os α / β hidrolase superfamília e superfamília Hotdog-fold. As duas superfamílias enzimáticas desempenham funções no metabolismo primário e secundário dos microrganismos e em eucariotos, as tioesterases estão presentes no citoplasma, retículo endoplasmático, mitocôndria (LATHAM, 2012). Além disso, as TEs são classificadas em Tipo I: contém um domínio α / β -hidrolase e as superfamílias carboxil-esterase e lipase, representam o maior grupo entre as tioesterases. Tipo II: conhecida como Hotdog-fold, apresenta seu sitio ativo com ácido aspártico ou ácido glutâmico (COHEN, 2013). Recentemente descrita a Tioesterase tipo III esta envolvida na via da β - oxidação de ácidos graxos (NIE et al.,2008).

A classificação das diferentes famílias e os nomes dos principais genes e substratos, e as famílias em negrito fora o foco do presente estudo (tabela 3),

Tabela 3. Famílias Tioesterase, nomes comuns dos seus membros e suas estruturas. Fonte: CANTU et al.,2010.

Família	Microrganismo Produtor	Estrutura	Gene/ nome família
TES 1	A,B,E ^a	NagB	Ach1
TES 2	A,B,E	α/β -Hidrolase	Acot1–Acot6, BAAT tioesterase
TES 3	A,B	Flavodoxin-like	<i>tesA</i> , acyl-CoA tioesterase I, protease I, lisofosfolipase L1
TES 4	B,E	HotDog	<i>tesB</i> , acyl-CoA tioesterase II, Acot8
TES 5	B	HotDog	<i>tesC</i> (<i>ybaW</i>), acyl-CoA thioesterase III
TES 6	A,B,E	HotDog	Acot7 (BACH), Acot11 (BFIT, Them1), Acot12 (CACH), YciA
TES 7	B,E	HotDog	Acot9, Acot10
TES 8	A,B,E	HotDog	Acot13 (Them2)
TES 9	B	HotDog	YbgC
TES 10	B	HotDog	4HBT-I
TES 11	B	HotDog	4HBT-II, EntH (YbdB)
TES 12	B,E	HotDog	DNHA-CoA hidrolase
TES 13	A,B	HotDog	paal, paaD
TES 14	B,E	HotDog	FatA, FatB
TES 15	B	HotDog	Thioesterase CalE7
TES 16	A,B,E	α/β -Hidrolase	TE domain of FAS (Tioesterase I), TE domain of PKS or NRP (type I Tioesterase (TE I))
TES 17	B	α/β -Hidrolase	TE domain of PKS
TES 18	B,E	α/β -Hidrolase	Tioesterase II, type II (TE II)
TES 19	B	α/β -Hidrolase	luxD
TES 20	E	α/β -Hidrolase	ppt1, ppt2, palmitoil-protein tioesterase
TES 21	A,B,E	α/β -Hidrolase	apt1, apt2, acil-protein tioesterase, fosfolipase, carboxyl esterase
TES 22	A,B,E	α/β -Hidrolase	S-formilglutathione hidrolase, esterase A, esterase D
TES 23	A,B,E	Lactamase	Hidroxi-glutathione hidrolase, glioxalase II

A, archaea; B, bactéria; E, eucarioto. Em negrito marca os produtores prevalentes.

2.4.3. Família Hot- Dog

A acil-CoA tioésterase ou Hotdog-fold é caracterizado por uma α -hélice central alongada "salsicha", envolta por cinco ou seis folhas- β antiparalelas "pão" (Figura 11) (LABONTE e TOWNSEND, 2012).

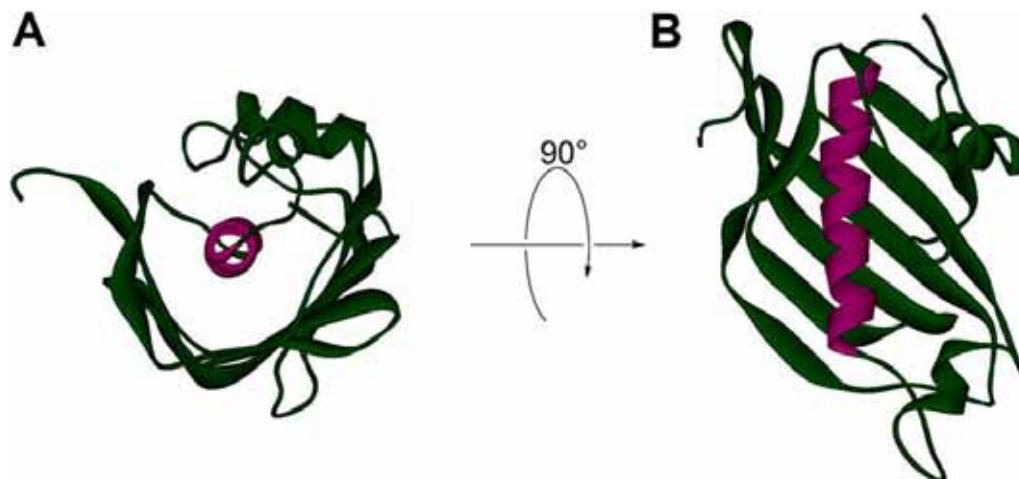


Figura 11. Hot-dog fold: A) *E. coli* desidratase-isomerase (FabA, PDB 1MKA) a α -hélice de cor ver roxa e folhas- β em verde, visto de frente. B) Giro em torno do eixo 90°. Fonte: LABONTE, TOWNSEND, 2012.

A unidade funcional mínima do membro do Hotdog-fold é um homôdimero e o sítio ativo está localizado na interface entre as duas subunidades, os resíduos catalíticos estão localizados no loop que se conecta com o N-terminal da α - hélice (DILLON e BATEMAN, 2002, PIDUGU et al., 2009; LATHAM, 2012).

As tioesterases Hotdog-fold possuem diferentes agrupamentos e conformações de estruturas quaternárias, representadas pelas abreviações D - dímeros dh - "hot-dog" duplo, TA - tetrâmeros TB - tetrâmeros com sobreposição de folhas β , H1 - hexâmero com sítio ativo na interface do loop, H2 - hexâmero com hélice na interface N-Terminal, H3- hexâmero com pareamento entre dímeros e Trimêros (Figura 12) (DILLON e BATEMAN, 2002, PIDUGU et al., 2009).

A família Hotdog-fold possui 1357 proteínas estruturalmente relacionadas e 17 subfamílias classificadas, e apesar do número de estruturas disponíveis, a função de muitas proteínas que contêm Hotdog-fold ainda são pouco conhecida (ANGELINI et al., 2008).

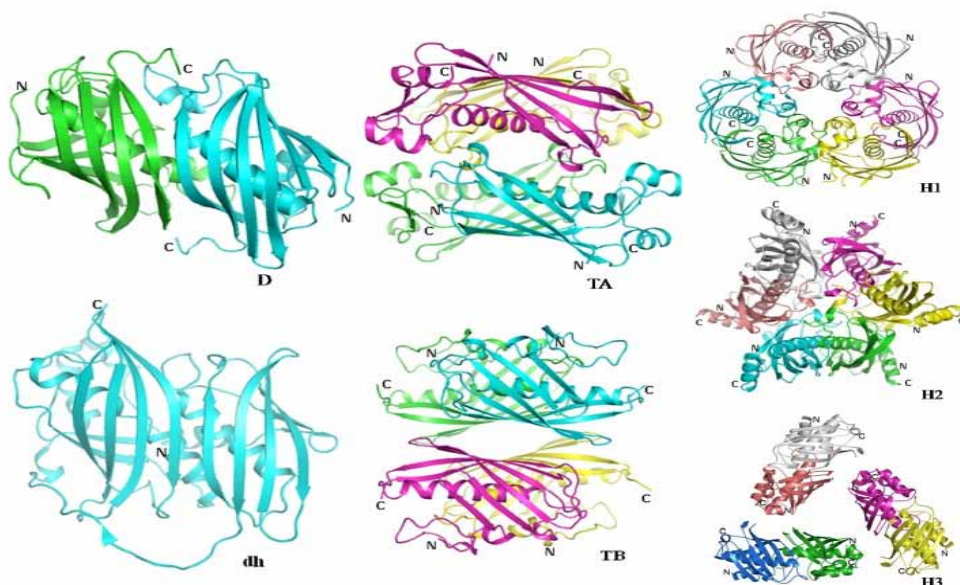


Figura 12. Diferentes associações da estrutura quaternária da proteína “hot-dog” (PIDUGU et al.2009).

Entre as diferentes subfamílias, as YbgC e YbaW, são de grande interesse científico e ainda pouco compreendidas (ANGELINI et al., 2008).

2.4.3.1. Subfamília YbgC

A proteína YbgC é codificada pelo primeira ORF do operon *tol-pal*, onde possivelmente participam na formação do anel de septos durante a divisão celular, mas a participação da proteína YbgC nesse processo não é clara. As YbgC está amplamente distribuída em bactérias gram-negativas, e possivelmente envolvidas na formação do anel de septos durante a divisão celular faz, além disso, apresentam preferências por substratos de cadeia curta (CANTU et al.,2010;PIDUGU et al., 2009)

Estudos recentes mostram que YbgC liga-se especificamente com o sistema de transporte de proteína acil (ACP) e está parcialmente envolvida com síntese de fosfolipídios e a biogênese de lipopolissacarídeos (LPS) na membrana celular (ANGELINI et al., 2008).

2.4.3.1. Subfamília YbaW

A Tioesterase codificada pelo gene *ybaW*, quebra as ligações tioéster da acetil-coenzima A (acil-CoA), por produtos gerados durante a β -oxidação de certos de ácidos graxos insaturados (NIE et al.,2008).

Ren et al (2004) usando *E. coli* como modelo verificaram que o cultivado em ácido oleico como fonte única de carbono, a referida bactéria obtém sua energia a partir da β -oxidação de ácidos graxos. Assim, foi proposto que noventa por cento do ácido oleico é oxidado à acetil-CoA na via da β -oxidação. Mas recentemente, Latham (2012) demonstrou que os dez por cento restante foi parcialmente oxidado por uma via alternativa, convertendo 2,5-trans-cis-tetradecadienoil-CoA em 3,5-cis-tetradecadienoil-CoA redutase (Figura 13). O produto secundário 3,5-cis-tetradecadienoyl-CoA não pode ser oxidado, devido ao fato de que a enzima necessária, dienoil-CoA-isomerase, não está presente.

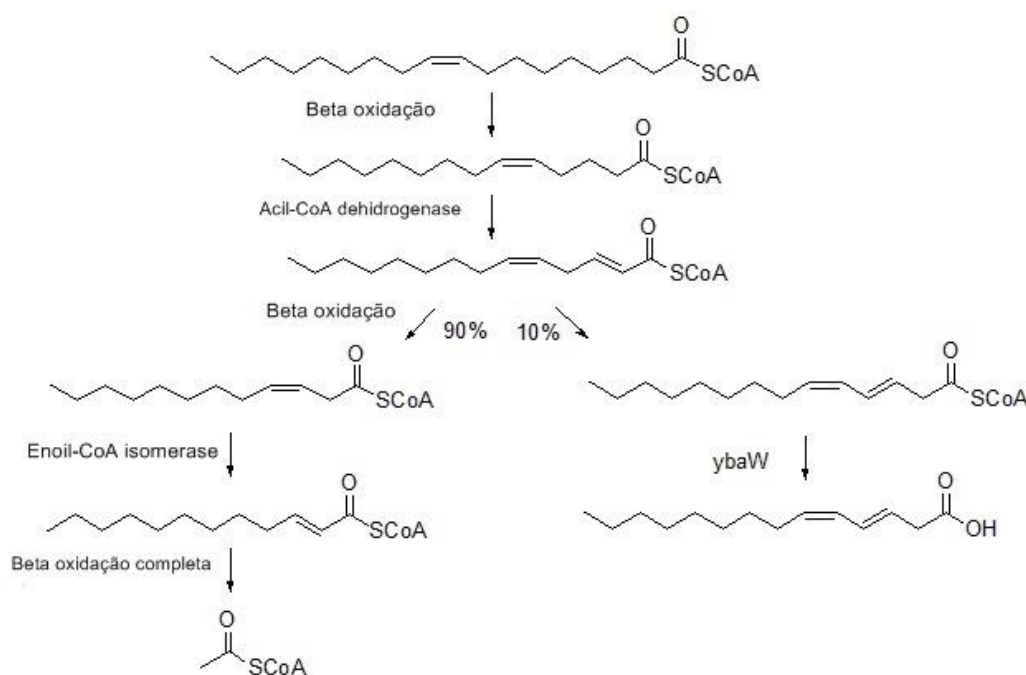


Figura 13. Beta-oxidação da *E. coli*. O valor de 90% refere-se à percentagem de oleoil-CoA que é degradado via Beta oxidação, ao passo que 10% restante foi degradado pela via alternativa, facilitando a degradação parcial de oleoil-CoA em ácido 3,5-tetradecadienóico. Fonte: LATHAM, 2012.

Foi realizado um estudo em *E. coli* com a mutação $YbaW^-$, e foi observado desenvolvimento perturbado ácido oleico, bem como o ácido linoleico conjugado, quando comparado com o desenvolvimento da *E. coli* $YbaW^+$ sob as mesmas condições (NIE, et al., 2008 a; NIE et al., 2008 b).

Analizando os resultados quais em conjunto, estes indicam que a tioesterase YbaW é necessária para a via alternativa β -oxidação. Outro fator que corrobora com essa evidencia é ao analisar o papel do gene *fadR* está localizado “upstream” do gene *ybaW*.

A proteína FadR tem um duplo papel, na *E. coli*, atuando como um repressor da via β -oxidação e como um ativador dos genes de biossíntese de ácidos gordos insaturados. Usando ensaios de mobilidade eletroforetica, Feng et. al., (2009) foi capaz de demonstrar que FadR se liga à região promotora *ybaW*, tornando YbaW o mais novo membro da *fad* regulon (responsável pelos genes codificadores das enzimas da β - oxidação) (Figura14).

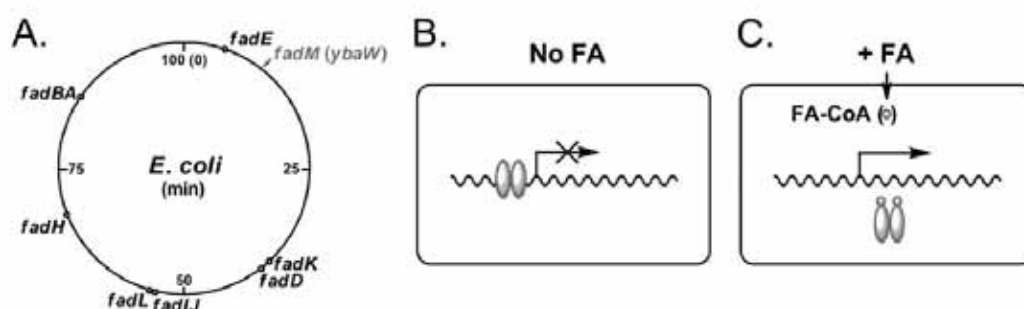


Figura 14. Regulação do gene *fad* e o modelo atual de sua regulação por ácidos graxos: A localização física do *fad* regulon na *E. coli*, o gene *ybaW* (*fadM*) está em cinza. B- Repressão do *fad* gene na ausência de ácidos graxos exógenos. C- Indução do *fad* gene após a adição de ácidos graxos exógenos. As estruturas ovais representam FadR, enquanto que os círculos indicam ligantes reguladores de acil-CoA e FA, ácidos graxos. Fonte: FENG e CRONAN, 2009.

3. OBJETIVO GERAL

Tendo em vista à constante busca por novas enzimas, o objetivo deste trabalho foi a seleção de proteínas de capacidade hidrolíticas como tioesterase e pirofosfatase inorgânica, a partir de uma sub-biblioteca metagenômica de solo de arboreto de eucalipto, com a perspectiva de obtenção de enzimas para utilização em processos industriais.

3.1. Objetivos específicos

- Anotação das sequências por meio da identificação de “Open Read Frames” (ORFs) e ProDom;
- Identificação dos genes tioesterase e pirofosfatase inorgânica;
- Predição da estrutural das proteínas com auxílio dos programas I-Tasser e PyMol;
- Desenho dos pares de oligonucleotídeos iniciadores para clonagem no vetor de expressão do sistema pET28.
- Testes de expressão proteica.

4. MATERIAL E MÉTODOS

4.1. Biblioteca e sub-biblioteca metagenômica

A biblioteca metagenômica utilizada para a prospecção de enzimas hidrolíticas foi construída com solo sob arboreto de *Eucalypto* spp., o DNA metagenômico foi clonado em vetor cosmídeo para a construção da biblioteca metagenômica (SCHUCH et al., 2009). Em estudo anterior SILVEIRA et al. 2006 comparou sequências metagenômicas de 16s rRNA de duas bibliotecas, utilizando solo da mesma região e outra com solo de floresta nativa. Os autores demonstraram que o solo sob arboreto de *Eucalypto* spp., possui maior diversidade bacteriana do que o solo da floresta nativa sugerindo assim que seria um bom indicador para descoberta de novas enzimas.

A prospecção de enzimas hidrolíticas e a construção da sub-biblioteca foram realizadas durante o trabalho de mestrado de Rodrigues (2008) que sequenciou 864 clones da sub-biblioteca com inserto de aproximadamente 30.000 pb, gerando 305 contigs, que selecionou o contig 241. Após a análise das sequências foi observado que o mesmo não estava completo. Para o fechamento e montagem foi construído um par de oligonucleotídeos iniciadores de extensão *primer walking* de acordo com a tabela 4.

Tabela 4. Iniciadores utilizados para fechamento do contig.

Iniciadores	Sequência 5'-3'
Gap_F	CGGCCCTCCGTACTC
Gap_R	TGCTGTGCCCTGCGAT

4.1.2. Extração DNA Cosmidial

A extração de DNA cosmidial foi feita pelo método de lise alcalina (SAMBROOK et al., 1989), para extração em tubos de 2 mL, o método foi realizado com algumas modificações. O clone estocado foi inoculado (10 µL) em 1 mL de meio LB, adicionado de 70 µg/mL de ampicilina, no meio

de cultivo de bactérias foi feito em duplicata. As amostras foram incubadas com agitação de 250 rpm, a 37°C, por 18 horas. Após a incubação, as amostras foram centrifugadas por 10 minutos, a 3220 $\times$ g, a 20°C. O sedimento foi lavado e ressuspendido em 60 μ L de solução GTE [Glicose 50 mM; Tris-HCl 23 mM (pH 8,0); EDTA 10 mM (pH 8,0)], com agitação em vortex por 2 minutos e centrifugação a 3220 $\times$ g a 20°C, por 10 minutos. Esse procedimento foi repetido mais uma vez. Após a lavagem acima descrita as células foram ressuspendidas em 60 μ L de solução de GTE contendo de 5 μ g de ribonuclease A (Sigma). Acrescentou-se 120 μ L de solução de lise [NaOH 0,2 N / SDS 1% (p/v)] a suspensão, e esta foi incubada por 10 minutos em temperatura ambiente. Após esse período foi adicionado 120 μ L de acetato de potássio 3M, repetindo-se o período de incubação anterior. Após esse período as amostras foram resfriadas a -80°C por 30 minutos, em seguida as amostras ficaram em temperatura ambiente para descongelar e foram centrifugadas a 3220 $\times$ g a 20°C por 45 minutos. O sobrenadante de cada amostra foi transferido para tubos de 2 mL estéreis, e adicionado 16 μ L de acetato de sódio a 3M pH 5.2 e 500 μ L de isopropanol absoluto (v/v), os tubos foram invertidos 10 vezes, para misturar as amostras e armazenadas a -20°C por proximamente 18 horas. Posteriormente as amostras foram centrifugadas a 3220 $\times$ g, por 45 minutos a -4°C. O sobrenadante foi descartado e o DNA precipitado foi lavado com 200 μ L de etanol 70% (v/v) gelado. Os tubos contendo as amostras de DNA foram centrifugados por 10 minutos nas mesmas condições de rotação e temperatura já descritas. Após esse procedimento, o sobrenadante foi descartado deixando os resíduos secarem por 60 minutos em temperatura ambiente. As amostras foram ressuspendidas em 30 μ L de água ultra pura Milli-Q (Millipore) esterilizada, e armazenada a -20°C.

4.1.3. Reação de sequenciamento

As reações de PCR, para o sequenciamento dos fragmentos de DNA foram realizadas nas seguintes condições: 2 μ L de BigDye Terminator (Applied Biosystems); 10 pmoles do oligonucleotídeo (GAP_F e R); 3 μ L

tampão 5X (400 mM Tris-HCl pH 9; 10 mM MgCl₂); 0,2 µg de DNA e água Milli-Q esterilizada para completar um volume de 10 µL. Foram feitas reações individuais para cada oligonucleotídeo utilizado. As placas foram seladas com um adaptador de silicone e levadas ao termociclador, sendo submetidas ao seguinte programa: 40 ciclos (96°C por 30 segundo, 52°C por 30 segundos e 60°C por 4 minutos). Após a reação de sequenciamento os fragmentos de DNA amplificados foram precipitados, e os dNTPs marcados por fluorescência que não foram incorporados, foram retirados por sucessivas lavagens. Para a precipitação do DNA amplificado, marcado pela PCR de sequenciamento, adicionou-se 80 µL de isopropanol 75% (v/v) às amostras, as placas foram agitadas, cuidadosamente, em vortex por alguns segundos, posteriormente incubadas à temperatura ambiente por 15 minutos, e centrifugadas a 20°C por 45 minutos, a 3220 xg. Os sobrenadantes foram descartados e 150 µL de etanol 70% (v/v) foram adicionados às amostras. As placas foram centrifugadas por 10 minutos, na mesma temperatura e força centrífuga descrita anteriormente, e os sobrenadantes foram descartados. As placas foram secas durante 30 minutos no fluxo laminar na ausência de luz. Para a aplicação no sequenciador, as amostras foram ressuspensas em 9 µL de Hi-Di Formamide (ABI Prism), submetidas à desnaturação por 5 minutos, a 96°C. O sequenciamento foi realizado em um aparelho de capilar modelo 3700 DNA Analyzer ABI Prism (Applied Biosystems).

4.1.4. Análise das Sequências

A qualidade das sequências geradas foi avaliada utilizando-se o programa "Sequencing Analysis 3.4", o qual gerou os eletroferogramas que foram submetidos à análise pelo "Phred/Phrap/Consed" (GORDON et al., 1998). A seleção das sequências adequadas foi realizada pelo programa Phred, o qual analisa a qualidade das sequências, visualizando graficamente pelo Consed (EWING et al., 1988; EWING & GREEN, 1998) e gerando arquivos no formato "fasta", na qual o nível de exigência mínima

foi de 250 bases com qualidade Phred acima de 20, auxiliado pelo programa “Contgen.pl”

4.1.5 Seleção das ORFs

Foram identificadas as Fases Abertas de Leitura (ORFs - software ORF finder:<http://www.ncbi.nlm.nih.gov/projects/gorf/>) das sequências metagenômica, foi feito uma triagem manual de prováveis sequências com domínios conservados com base na maior similaridade em bancos públicos de dados integrados ao NCBI (BlastP, ALTSCHUL et al., 1990), seguido de correção de erros da anotação em códons de iniciação. As mesmas sequências nucleotídicas conservadas (ORFs) foram submetidas a um programa de identificação de domínios conservados: o PRODOM (SERVANT et al., 2002), buscando confirmação da predição funcional das sequências. As ORFs também analisadas às no programa Pfam (Banco de famílias enzimáticas: PUNTA et al., 2012) obtendo-se maiores detalhes sobre as sequências.

4.2. Identificações das proteínas hidrolíticas e análise estrutural

Após a seleção das ORFs foi feita a identificação de proteínas hidrolíticas, as sequências de amino ácidos foram submetidas no programa I-Tasser Server, plataforma on-line para predição de estrutura e função que permite gerar automaticamente predição de estrutura 3D e possível função biológica da proteína de alta qualidade (ZHANG, 2008).

Apos a predição estrutural as seqüências de aminoácidos de ambas foram analisadas no programa Swiss – Model (BORDOLI et al., 2009) para identificação de modelos homólogos as estruturas proteicas. A comparação entre a estrutura molde e o modelo foi feita com auxilio do programa PyMOL Molecular Graphics System, version 1.5.0.4, que é um sistema de visualização molecular. (SCHRÖDINGER, 2010). O programa RAMPAGE (LOVELL et al., 2003) permite gerar o gráfico Ramachandran atualizado, define os resíduos que se encontram nas regiões energeticamente

favoráveis e desfavoráveis e orienta a avaliação da qualidade de modelos teóricos ou experimentais de proteínas. O Programa PROSA WEB (WIEDERSTEIN e SIPPL, 2007) foi desenvolvido principalmente para avaliar estruturas experimentais, e calcula a significância estatística de uma estrutura por meio da comparação com estruturas já resolvidas por dois diferentes métodos, cristalografia de raios X e ressonância magnética nuclear (NMR).

O programa MEMSAT3 e MEMSAT-SVM possui método de identificação de topologia transmembranar (NUGENT et al., 2011). Para predição do motivo conservado foi utilizado o programa PROSITE (<http://prosite.expasy.org/>) (SIGRIST et al., 2010). O programa ESPript 2.2 (<http://esprict.ibcp.fr>) (GOUET et al., 1999) permite a visualização rápida do alinhamento de sequências além de avaliar estrutura secundária criados por para produzir uma síntese de ambas sequência e informações estruturais.

4.2.1 Análise Fenética

Preliminarmente, as sequências de aminoácidos foram submetidas à consulta de similaridade, com sequências depositadas no banco de dados GenBank acessado através do site do *National Center for Biotechnology Information* (NCBI). A ferramenta utilizada para esta consulta foi o BlastX local - *Basic Local Alignment Search Tools* (ALTSCHUL et al., 1997). Em seguida foi realizado o alinhamento global entre as sequências de aminoácidos selecionadas para as duas proteínas hidrolíticas, usando o programa BioEdit, version 5.0.9 (HALL, 2001). Para matriz de distância das árvores utilizou-se o método de construção “neighbor-joining” (SAITOU e NEI, 1987) com algoritmo p-distance respectivamente, processado pelo programa de Análise Genética de Evolução Molecular MEGA5 utilizando Bootstrap com 1.000 replicas. (TAMURA et al., 2011).

4.3. Amplificação e Clonagem da proteína em vetor expressão

Para cada gene foi construído um par de oligonucleotídeos iniciadores específicos contendo sítios para as enzimas de restrição (Tabela 5).

Tabela 5. Iniciadores para utilizados na amplificação dos genes codificadores de hidrolases.

Gene	Sequência (5'-3')	Sítio de restrição
<i>PPiase_F</i>	GATTATCAC <u>G</u> GATCCAAAATGGCGAAG	BamHI
<i>PPiase_R</i>	GCTTCTTG <u>G</u> AGCTCCTCAGGTCTTGTC	SACI
<i>TEs_F</i>	GGCGATCCGTCGATGGTCGCGT	BamHI
<i>TEs_R</i>	GTACAAGCTTCGGCCCTAGCGCTCG	HindIII

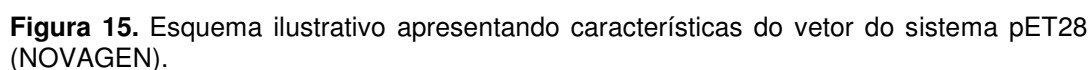
*Não foi obtido amplificação para o gene da pirofosfatase inorgânica (PPiase).

A sequência codificadora de tioesterase foi amplificada por PCR (Reação em cadeia da Polimerase) em um sistema contendo 1 U *Pfu* DNA Polymerase (Fermentas), 10 pmol de cada iniciador descrito acima, 200 mM de dNTP, 5 µL de tampão de reação (200 mM Tris-HCl 100 mM (NH₄)₂SO₄, 100 mM KCl e 20 mM MgSO₄) 10X concentrado, aproximadamente 50 ng de DNA molde (cosmídeo) e água ultra pura para completar o volume total de 50 µL.

A reação foi realizada em termociclador Eppendorf Master Cycler Gradiente utilizando-se o seguinte programa: 1 ciclo de 95° C por 2 minutos seguidos de 30 ciclos sequêciais de 95° C 45 segundos, 57° C 45 segundos e 72° C 2 minutos finalizando com um ciclo de 72°C por 5 minutos.

O produto de PCR foi purificado em coluna de sílica utilizando-se o kit *GFX PCR DNA and Gel Band Purification* (GE Healthcare). Em seguida, o DNA foi digerido com as enzimas de restrição especifica no desenho dos oligonucleotídeos iniciadores. As endonucleases foram inativadas por calor 20 minutos a 80° C e o DNA amplificado foi novamente purificado utilizando kit *GFX PCR DNA and Gel Band Purification* (GE). Por fim, o produto de PCR foi ligado no vetor de expressão pET28b digerido com as mesmas enzimas e purificado da mesma maneira (Figura 15). A reação de ligação

foi feita utilizando-se 10 ng de vetor, aproximadamente 50 ng do produto de PCR, 1 U de enzima T4 DNA ligase (Fermentas), tampão de reação diluído 10x (Tris-HCl 400 mM, $MgCl_2$ 100 mM, DTT 100 mM, ATP 5 mM) e água ultra pura estéril para um volume final de 10 μ L. A reação foi incubada a 16° C por 18 horas. Os plasmídeos foram transformados em *E. coli* BL21 (DE3), conforme instruções do fabricante (Novagen).



4.4. Expressão e purificação de proteína

4.4.1. Eletroforese em gel de poliacrilamida em condição desnaturante (SDS-PAGE)

SDS-PAGE é capaz de separar proteínas de acordo com sua massa molecular. A desnaturação é parcialmente promovida pela ligação do dodecil sulfato de sódio (SDS) à proteína formando um complexo proteína-SDS de carga negativa. As análises eletroforéticas de proteínas em gel SDS-PAGE foram realizadas utilizando o sistema de mini géis Mini-PROTEAN da BioRad. A composição dos géis utilizados é apresentada na tabela 6. Para a aplicação no gel as amostras de proteínas foram misturadas com tampão de amostra (Tris-HCl pH 6,8 100 mM, glicerol 20 % (m/v), SDS 4 % (m/v), azul de bromotimol 0,2 % (m/v) e beta-mercaptoetanol 2mM) na proporção 1:1 e aquecidas a 95° C por 5 minutos. As eletroforeses foram conduzidas a 180 V durante 50-60 minutos em tampão Tris-base 25 mM, glicina 250 mM e SDS 0,1 %. Após a corrida, o gel foi corado com Coomassie Brilhante Blue R-250 2,5 % (USB Corporation) durante 30 minutos. A seguir, o gel foi lavado em água ultra pura e descorado em solução descorante (água/metanol/ácido acético nas porcentagens 4,5/4,5/1,0 v/v/v).

Tabela 6. Composição do gel SDS-PAGE utilizado para análise de proteínas.

Componentes	Gel de empilhamento 5%	Gel de separação 12%
Tris-HCl pH 6,8 0,5M	1,0mL	-
Tris-HCl pH 8,8 2,25M	-	2,83 mL
*SDS 4%	100 µL	280µL
Bis acrilamida 30%	665µL	4,38mL
*APS 10%	50µL	150 µL
*TEMED	1,25 µL	25 µL
Água ultra pura	2,21mL	5,12mL

* APS - persulfato de amônio e TEMED - Tetrametil – etilendiamina. Valores equivalente para fazer um gel

4.1.2. Expressão de proteínas

Os testes de expressão e purificação foram feitos apenas para a proteína da tioesterase.

O transformante foi inoculado em meio LB pH 7,0 com 100 mg / mL de kanamicina, o pré-cultivo foi multiplicado a 37° C a 180 rpm até uma D.O.₆₀₀ 0,6. Nesse ponto foi retirado uma alíquota do pré-cultivo e transferido para um novo frasco com 2 L de meio LB pH 7,0 com 100 mg / mL de Kanamicina, e novamente colocado sob agitação 37° C a 180 rpm até uma D.O.₆₀₀ 0,4, nesse ponto o meio foi transferido para temperatura 18° C e a 280 rpm até atingir D.O.₆₀₀ 0,6 e nesse ponto foi adicionado ao meio de cultivo o indutor de expressão IPTG (Isopropil β-D-1-tiogalactopiranosideo) na concentração 1,0 mM (RODRÍGUEZ-GUILBE et al., 2013). Para o teste de indução a cultura ficou o período de 22 horas a 18° C e a 280 rpm. Em seguida as células do cultivo foram centrifugadas a 3220 *xg* / 10 min., o precipitado foi ressuspensionado com 5 mL de Fosfato de Sódio 0, 1M pH 6,5 e 0,5 mg / mL lisozima e encubado em gelo por 30 min. O choque térmico foi realizado mantendo o extrato por 30 min. a -80° C e após esse período, levado imediatamente em banho a 37° C. O processo do choque térmico foi feito por três vezes, a cada etapa as amostras foram centrifugadas a 3220 *xg* por 30 min. a 4° C e o sobrenadante foi recuperado a cada centrifugação. A quantidade de proteínas nas frações solúvel e insolúvel foi avaliada em gel SDS-PAGE

4.1.3. Purificação da proteína

O extrato bruto (7 mL) foi concentrado por meio de filtração em membrana Thermo Scientific Pierce® Protein Concentrators, 7mL/20K MWCO, e a fração proteica que ficou retida foi aplicada em uma coluna cromatográfica de níquel pré-empacotada *HisTrap™FF crude* 1mL and 5 mL (GE Healthcare) e a purificação realizada de acordo com as instruções do fabricante.

5. RESULTADOS E DISCUSSÃO

5.1. Biblioteca, sub-biblioteca metagenômica e montagem do contig

O inserto de DNA presente no cosmídeo A09 possui 30.562 pares de bases, a busca por regiões codificadoras foi feita através dos programas ORF Finder, onde foram selecionadas 24 ORFs (Open Reading Frames). A montagem do mapa gênico pode ser visualizada na figura 16. As sequências selecionadas foram submetidas no GenBank, número de acesso da KF715620, e estão descritas na tabela A1.

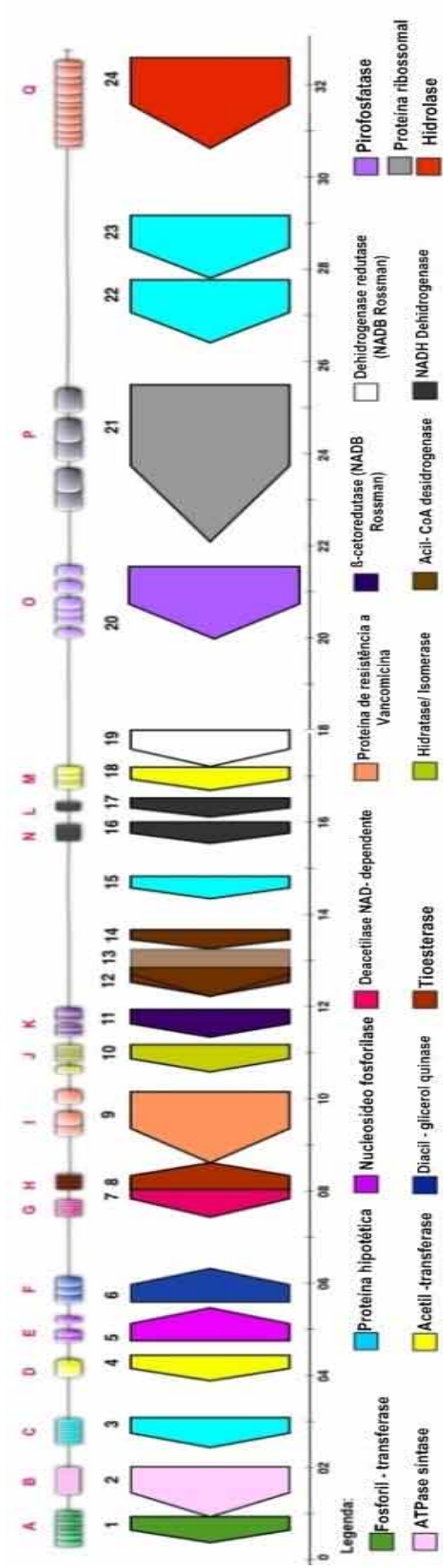


Figura 16. Organização funcional do cosmídeo: As setas indicam o tamanho relativo e a direção da transcrição de cada ORF.

5.2. Identificações das proteínas hidrolíticas e análise das estruturas

5.2.1. Identificações da Pirofosfatase Inorgânica (PPIase) e análise das estruturas

Após predição da estrutura da Orf 20 utilizando o programa I-Tasser, as sequências homologas foram identificadas utilizando o programa Swiss-Model (Tabela 7) (ARNOLD et al., 2006), visando selecionar, preferencialmente, moldes que apresentaram identidade igual ou superior à 30%, haja visto que este valor representa conservação confiável de resíduos de aminoácidos das cadeias peptídicas (SANTOS FILHO e BICCA DE ALENCASTRO, 2003).

A Orf 20 foi identificada como sendo uma pirofosfatase inorgânica (PPIase), e para melhor entendimento será denominada de MetaPPIase.

Tabela 7. Predição dos modelos para MetaPPIase

Código PDB	Função	Cobertura	ID %	Método	Formação estrutural	Ligantes
4a01.1	Proton	0,79	25,7	X-ray,	Homôdimero	DMU*, MG, K, PN
<i>Vigna radiata</i>	Pirofosfatase		5	2.3Å		
4av6.1	Pirofosfatase	0,69	32.1	X-ray,	Homôdimero	MG, K
<i>Thermotoga maritima</i>	Bomba de sódio Energizada K ⁽⁺⁾		2	4.0Å		
4av3.1	Pirofosfatase	0,77	26.5	X-ray,	Homôdimero	MG, CA
<i>T. maritima</i>	Bomba de sódio Energizada K ⁽⁺⁾		5	2.6Å		

* DMU beta - decil-D-Maltopiranoside, MG íon magnésio, K íon potássio, PN ácido Imidodifosforico, CA íon cálcio.

Observa-se na tabela 7 que as sequências homologas identificadas pelo programa Swiss-Model (ARNOLD et al., 2006), sequência em destaque, representa o molde com identidade acima de 30% que pertence *Thermotoga maritima*, sendo o número de acesso do PDB 4av6. Com a identificação do molde foi possível prever o modelo estrutural da proteína MetaPPIase, utilizando o programa Pymol (Figura 17).

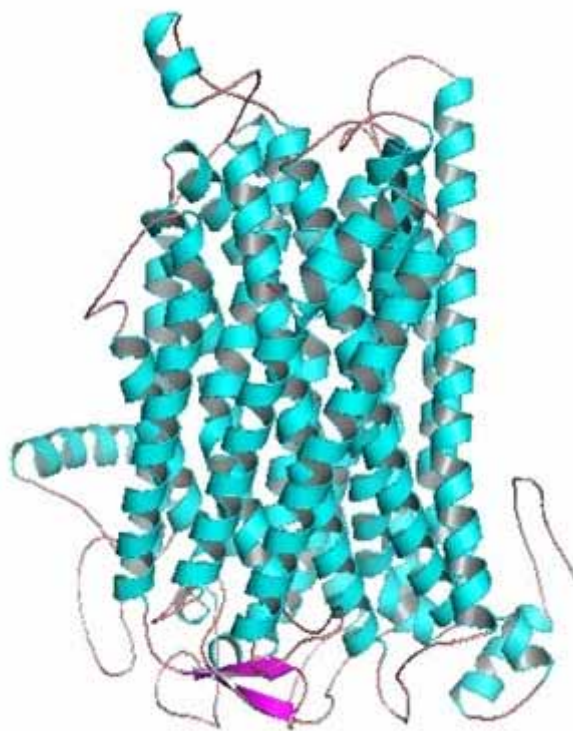


Figura 17. Predição estrutural para MetaPPiase: onde as α -hélices estão representadas pela cor ciana, folhas- β em pink e os loops salmão. A estrutura foi criada no programa Pymol version 1.5.0.4

A confiabilidade do modelo gerado para MetaPPiase foi avaliada através do gráfico de *Ramachandran*, que determina que 90% dos resíduos aminoácidos devem estar localizados na região mais favorável do gráfico (LOVELL, 2003). A análise gerada para MetaPPiase obteve 91,2 % de aminoácidos localizados na região mais favorável, representada pela área azul escura do gráfico, é importante observar na figura 16 que apresenta maior quantidade de aminoácidos na região referente a α -hélice de sentido horário, já que a estrutura da MetaPPiase na figura 18 é formada basicamente por estruturas α -hélices. Uma parte considerável dos resíduos de aminoácidos 5,6 % ficou na região favorável glicina, mas por se tratar de uma região flexível acaba não interferindo na estrutura. E apenas 3,2 % dos resíduos ficaram na região não permitida.

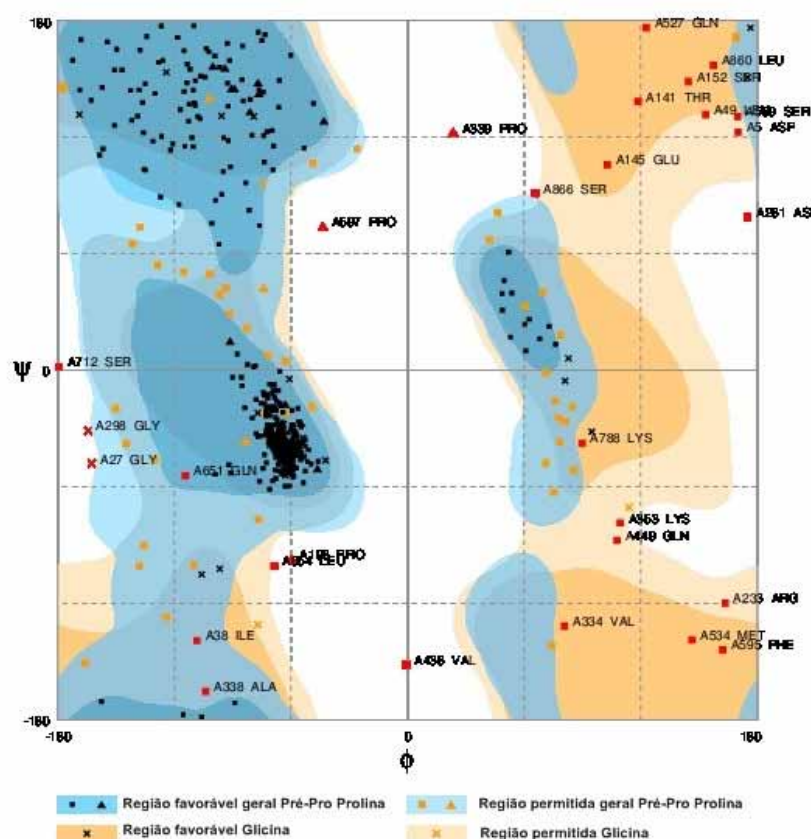


Figura 18. Gráfico Ramachandran baseado nos ângulos de torção ϕ e ψ : a região azul escura representa região favorável geral Pré-Pro Prolina a azul clara região permitida geral Pré-Pro Prolina; em laranja escuro região favorável geral para Glicina e laranja claro região permitida geral para Glicina. Gráfico foi gerado no programa Rampage.

Outro fator importante avaliado foi o valor de z-score com auxílio do programa ProSA web é uma ferramenta de diagnóstico que se baseia na análise estatística das estruturas de proteínas disponíveis (WIEDERSTEIN e SIPPL, 2007).

O z-score para MetaPPIase foi de $-7,09$ (Figura 19), de acordo com a descrição do programa o valor negativo pode ser pelo tamanho da proteína, por ser uma estrutura complexa acima de 800 resíduos de aminoácidos tornando difícil a modelagem experimental, outro fator que induz pontuação negativa do z-score esta relacionado com estruturas das proteínas de membrana, pois suas propriedades físico-químicas diferem consideravelmente de proteínas solúveis, no entanto muitas estruturas publicadas em bancos de dados estão com sequências incorretas (BENKERT et al., 2011; WIEDERSTEIN e SIPPL, 2007).

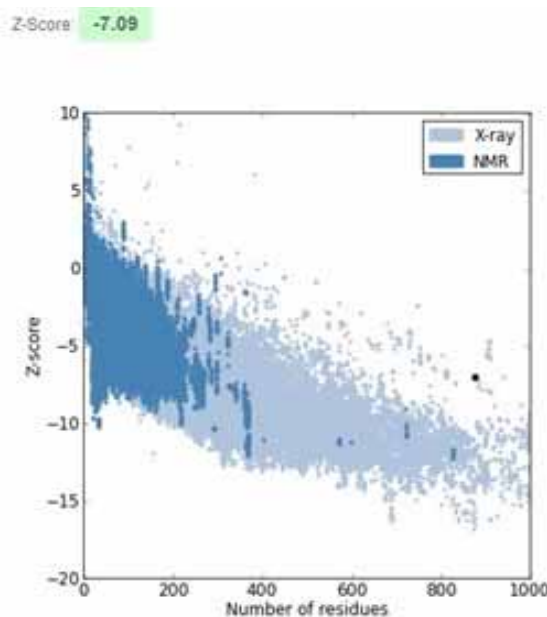


Figura 19. Pontuação de Z-score: pontos indicados com azul escuro são estruturas de proteína resolvidas por ressonância magnética nuclear (NMR) e pontos em azul claro corresponde as estruturas resolvidas através cristalografia de raios-X, o círculo preto mostra no gráfico o local da pontuação para MetaPPIase (Prosa Web).

De acordo com a análise estrutural e de confiabilidade, podemos dizer que o modelo gerado para MetaPPIase é confiável baseado nas pontuações do gráfico de *Ramachandran*, mesmo o modelo apresentando 3,2% de resíduos que estavam em região isolada, já que se trata de uma estrutura com 876 aa. O valor negativo de z-score indicado para MetaPPIase pode ser aceito de por se tratar de uma proteína de membrana e ainda não existe programas específicos para análise de estruturas proteínas de membrana.

A sequência homologa à MetaPPIase pertence a *T. maritima* (PDB 4A06) a proteína pertence à família pirofosfatase integral de membrana (M-PPIase), e a principal característica dessa família é apresentar entre 15 a 17 hélices transmembrânicas (TM) (KAJANDER et al., 2013). Característica essa confirmada pelo programa MEMSAT3 e MEMSAT-SVM (Figura 20).

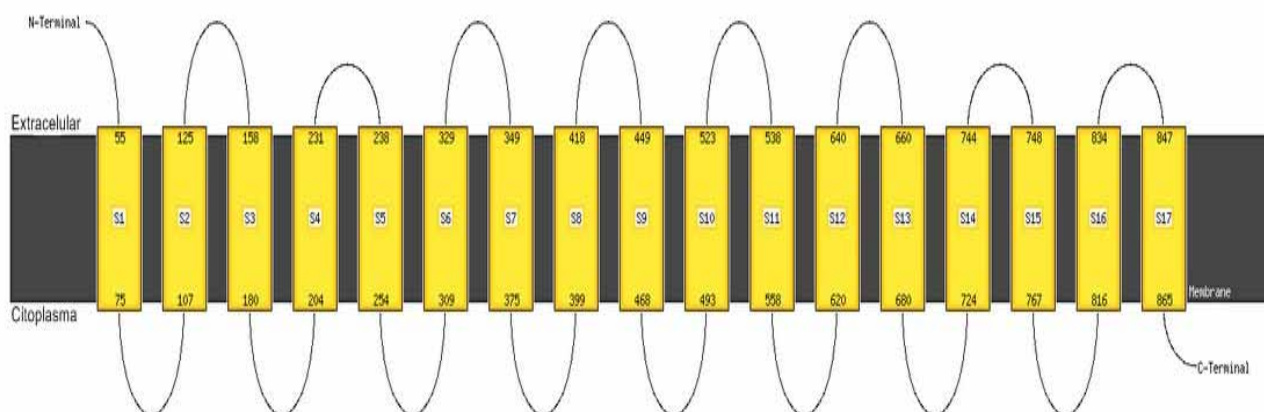


Figura 20. Topologia transmembranar: Modelo dos segmentos transmembranares da MetaPPIase.

Segundo, RODRIGUES et al. (2014), a MetaPPIase possui 17 hélices TM representadas pelos segmentos amarelos, a porção C-terminal voltada para região citoplasmática e a N-terminal voltada para região extracelular além de peptídeo sinal (Figura 20). Essas características são similares às descritas para *Streptomyces coelicolor* que é classificada como uma M-PPIase ou pirofosfatase transportador de H^+ ou (H^+ -PPIase) (MIAMURA et al., 2004). As H^+ - PPIase são divididas em duas subfamílias a K^+ independente e K^+ dependente, sendo as K^+ independentes identificadas em eucariotos, bactérias e archaea, enquanto as K^+ dependentes identificadas em eucariotos e em algumas plantas (LUOTO et al., 2011; BAYKOV et al., 2013, Wang et al., 2013). Recentemente foi identificado uma pirofosfatase transportadora de Na^+ (Na^+ - PPIase) em microrganismo mesófilos, e esta enzima é semelhante a H^+ -PPIase em muitos aspectos e para a atividade requerer tanto K^+ e Na^+ para realizar transporte de íons através da membrana (KAJANDER et al., 2013; BAYKOV et al., 2013).

Além da presença das hélices transmembranares, a predição do motivo conservado da proteína MetaPPIase, D-[SGDN]-D-[PE]-[LIVMF]-D-[LIVMGAC] foi identificado utilizando o programa PROSITE (SIGRIST et al., 2010). LIN et al. (2012), descreveram o motivo (DX3DX3D) como responsável pela função ácida, devido ao fato de que os três resíduos de aspartato conservados serem característicos da família da H^+ -PPIase.

Wang et al., (2013) caracterizaram diferentes domínios conservados pertencentes à família H⁺-PPiase de plantas usando diversas ferramentas de bioinformática, que permitiu a identificação de dois possíveis novos domínios os quais permitem alocar a estrutura e função entre os membros da família H⁺-PPiase.

Para observar o domínio conservado da proteína MetaPPiase e os diferentes organismos representantes da H⁺PPiase, foram selecionadas sequências de aminoácidos dos seguintes microrganismos: *Akkermansia muciniphila* Am-PPiase, *Clostridium leptum* Cl- PPiase, *Clostridium lentocellum* Clem-PPiase, *Prevotella oralis* Po-PPiase, *Vignia radiata* Vr-PPiase, *Thermotoga maritima* Tm-PPiase, *Bacterioides vulgatus* Bv-PPiase e *Clostridium sp.* Cs-PPiase. O alinhamento das sequências foi feito utilizando o programa ESPript 2.2 (GOUET et al., 1999) que agrupa e alinha sequência e a estrutura secundária para as análises (Figura 21).

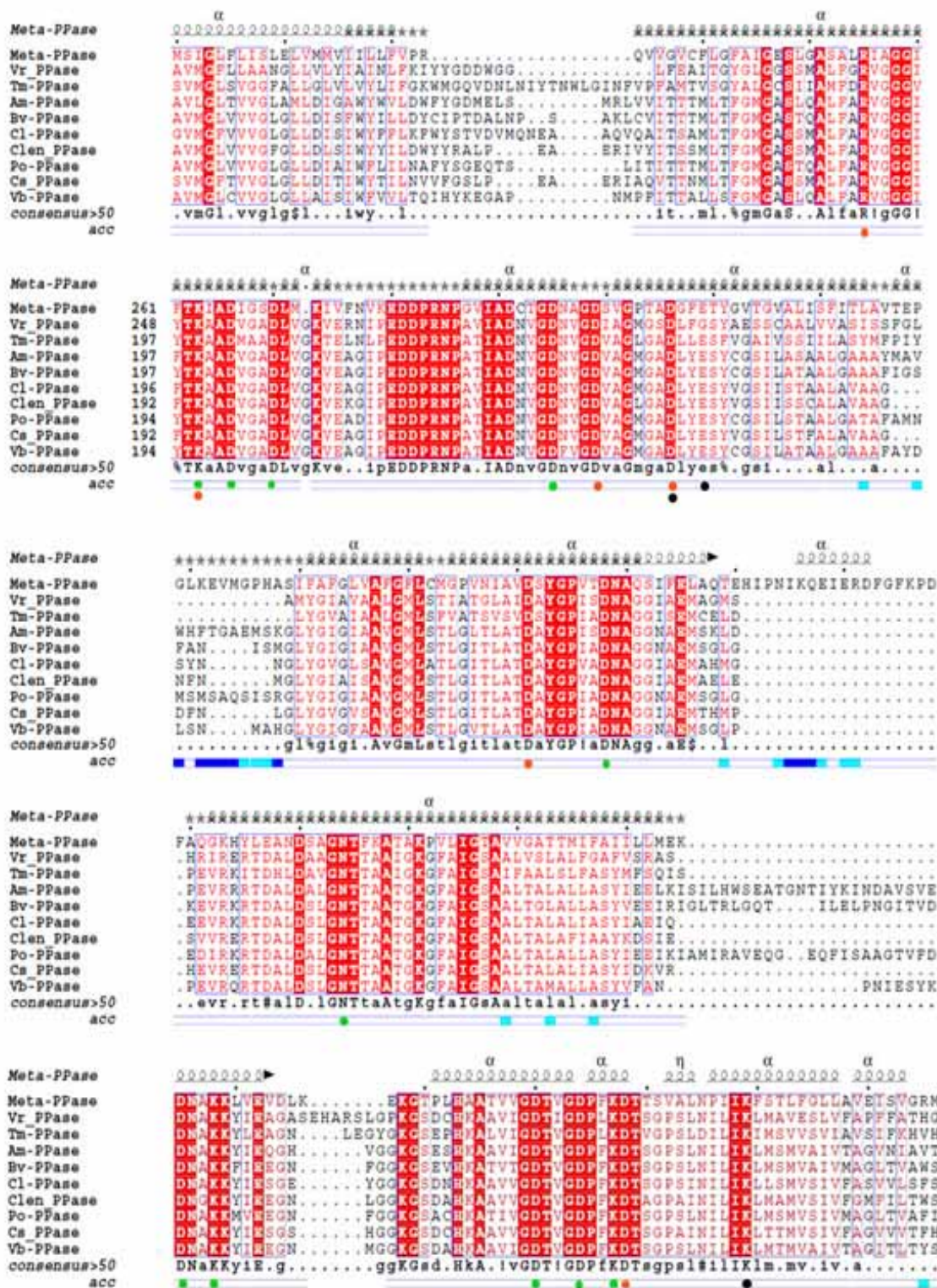


Figura 21. Alinhamento múltiplo das sequências da H⁺PPase: Resíduos totalmente conservados estão marcados em vermelho, e regiões com resíduos homólogos estão em marcados na caixa azul. Os resíduos Mg²⁺ e PPI *binding* estão marcados com esferas verdes; resíduos do túnel que estão ligados no local de hidrólise e o canal de íons, indicado com esferas laranja; resíduos de abertura do canal marcado com esferas pretas e a (acc) regiões variáveis da entre as sequências representado pela cor azul.

O alinhamento da figura 21 revelou elevado grau de homologia entre as diversas sequências de aminoácidos. Nas regiões em vermelho, estão os ligantes Mg^{+} , K^{+} e o PPI, responsáveis pela formação do canal transportador de íons (Figura 19) (LIN et al., 2012).

Para melhor visualização dos motivos conservados indicados na figura 21, foi feita sobreposição das estruturas MetaPPIase com o modelo Tm-PPIase, representados na figura 22, para isso foram usadas as coordenadas dos ligantes do modelo da estrutura descrita por Kellosalo et al. (2012).

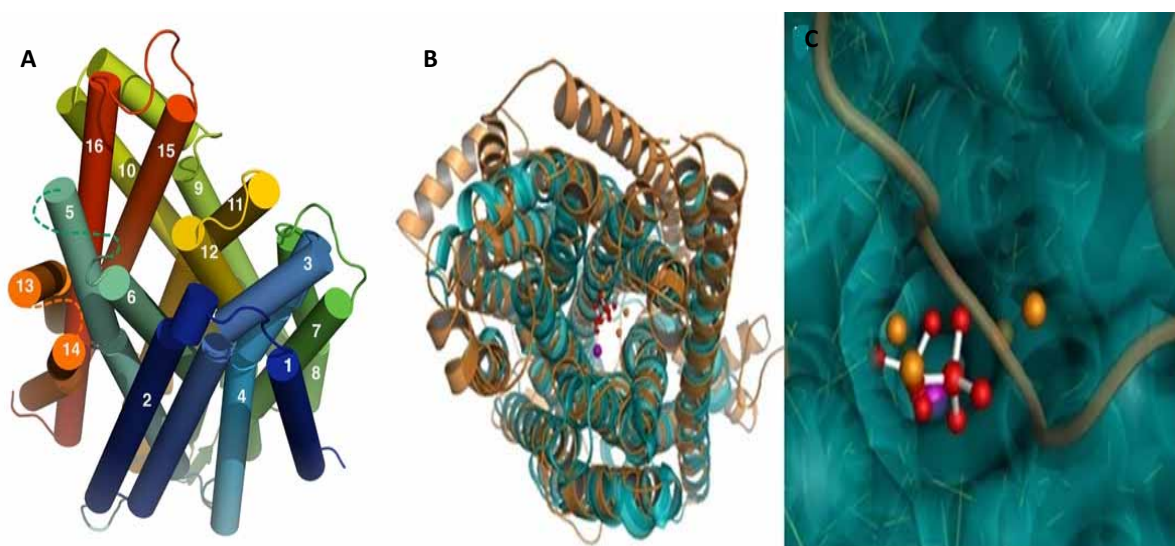


Figura 22. Possível coordenada do substrato para MetaPPIase: A) Vista inclinada de um monômero Tm-PPIase do lado citoplasmático, a vista enfatiza as posições das estruturas transmembranares desordenadas, que estão numeradas de acordo com a gradiente de cores (azul para vermelho) do N para o C terminal. Fonte: Kellosalo et al 2012. B) Sobreposição das estruturas MetaPPIase em azul e Tm-PPIase laranja clara estrutura na cor laranja claro, é possível ver os ligantes de uma visão geral onde fosforo está em vermelho, Mg^{2+} laranja e K^{+} roxo. B) Visão aproximada do túnel. Fonte: As figuras B e C foram desenhadas no programa Pymol version 1.5.0.4.

O valor do desvio mínimos quadrados (rmsd) para sobreposição das estruturas foi de 2.322 Å, valor esse próximo do ideal que é da ordem de 2 Å. A figura 22 B mostra a sobreposição das estruturas juntamente com os ligantes formadores do canal transportador de íons e podemos dizer que possivelmente a MetaPPIase pode atuar como uma bomba de prótons reversível. Kellosalo et al. (2012), em seus estudos para entenderem a estrutura da M-PPIase presente em *Thermotoga marítima* identificaram os

domínios transmembranares TM 5 e 6 e TM 11 e 12 como sendo as estruturas responsáveis pela abertura e fechamento do canal da como uma bomba de prótons reversível. Esse sistema pode ser melhor compreendido observando a figura 23.

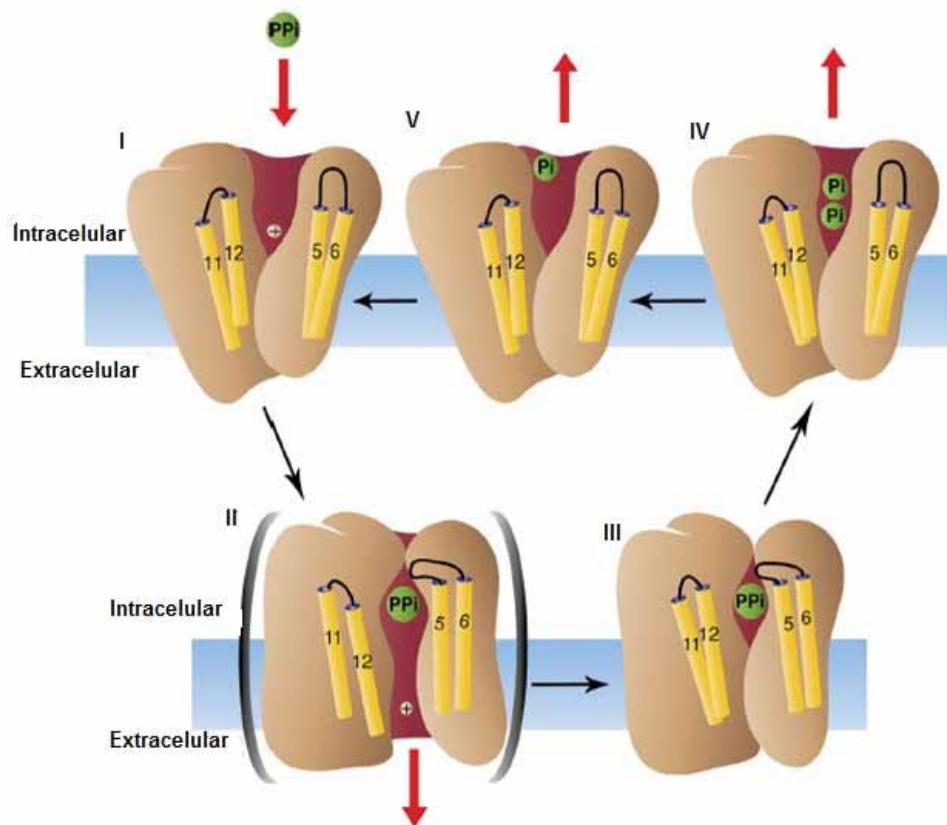


Figura 23. Mecanismo proposto para a bomba de prótons reversível de *Thermotoga maritima*: A faixa azul simboliza a membrana celular; a numeração de I a V indicam a ordem dos acontecimentos. I) Antes da ligação do substrato é possível observar o canal para a região extracelular fechado e a região intracelular aberta (sítio ativo). II) A ligação do substrato faz com que o circuito entre hélices 5 e 6 mude sua conformação, fechando a região intracelular, permitindo a formação de estado transitório onde o movimento das hélices 11 e 12 abra a região extracelular (canal de saída), e ocorre o transporte de um cátion para meio extracelular. III) O sítio ativo mantém fechado até a fim da hidrólise do pirofosfato. IV, V) Após hidrólise o grupo fosfato as duas molecular de fosfato inorgânico são levados para dentro da célula. E o canal volta para conformação inicial para uma nova hidrólise. Fonte: Kellosalo et al.,2012.

5.2.2. Fenograma

A construção do fenograma foi feito através da seleção de sequências aminoácidos representativas da família H^+PPi ase depositadas no NCBI, essas sequências estão listadas na tabela A2 no apêndice.

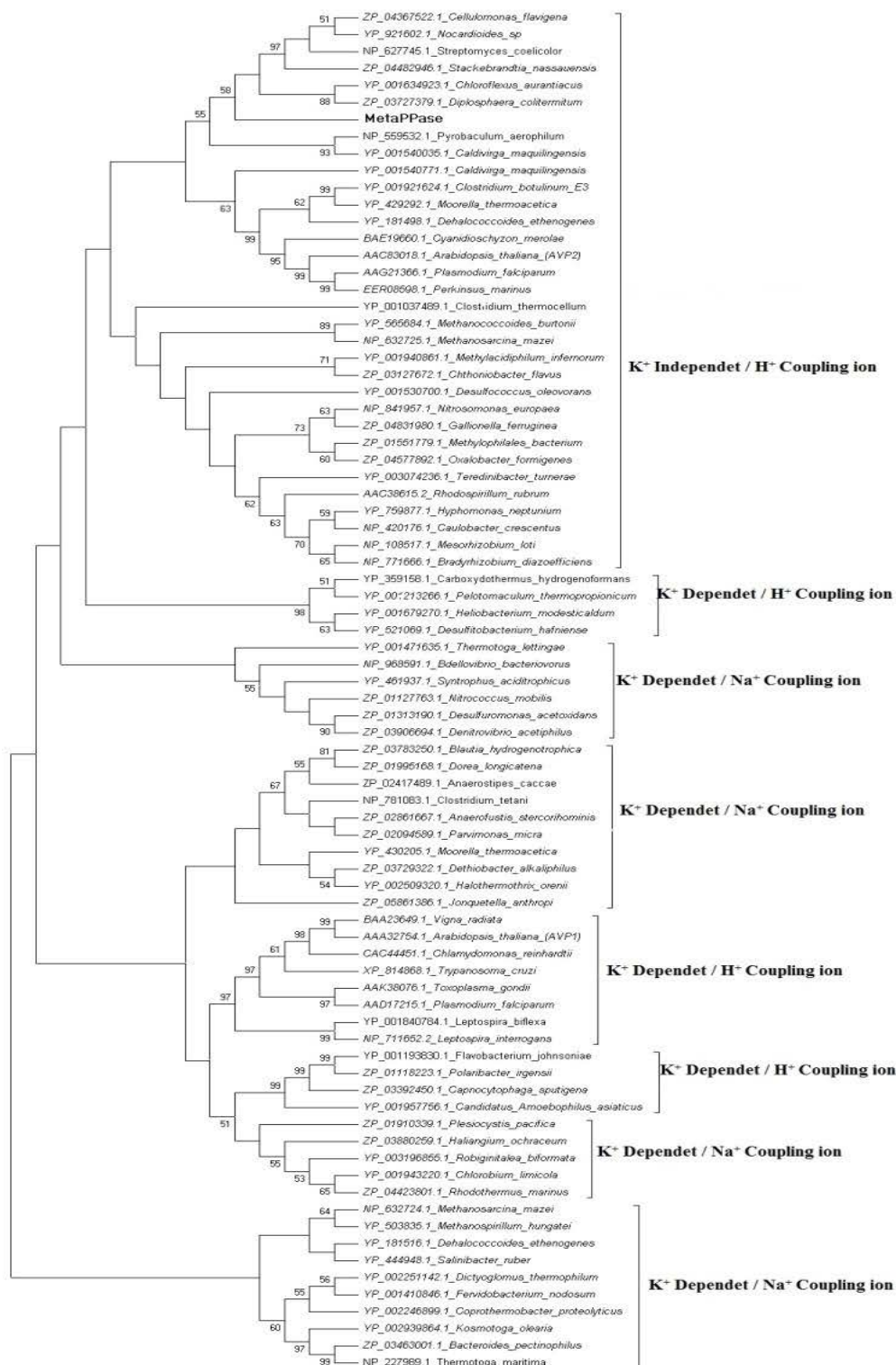


Figura 24. Análise Fenética MetaPPiase: As sequências de membrana PPIase foram selecionadas de acordo com o banco de dados NCBI para cada subfamília PPIase (H⁺-PPIase com K⁺ independente e K⁺ dependente).

De acordo com as informações obtidas para a MetaPPiase, sugere que ela pertence à família -PPiase ou H⁺ - PPiase, que são divididas em duas subfamílias a K⁺ independente e K⁺ dependente.

O fenograma mostra que a MetaPPiase pertence ao grupo de sequências da subfamília K⁺ independente com afinidade para íon H⁺, encontrada na maioria dos microrganismos, principalmente, bactérias (LUOTO et al., 2011; BAYKOV et al., 2013, Wang et al., 2013). Observa-se ainda que MetaPPiase está no limite entre dois clados demonstrando a singularidade dessa proteína (Figura 24).

5.2.3 Identificações do gene Toiesterase (TEs) e análise das estruturas

Para a predição da ORF 08 foram realizadas as mesmas descritas no item 5.2.1., e foi identificada como sendo uma Tioesterase (TEs), e para melhor entendimento será chamada de TEs-Meta.

A tabela 8, mostra as sequências homologas da proteína TEs-Meta, obtidos pela análise do programa Swiss - Model (ARNOLD et al. 2006), para escolha da sequência molde foram usados os mesmo parâmetros do item 5.2.1 de no mínimo 30% de identidade, a sequência em negrito na tabela 8 foi que apresentou maior identidade (*Thermus thermophilus* número de acesso PDB 2cye) .

É importante destacar que dentre as estruturas descritas na tabela 8, apenas três possuem artigos publicados, o restante são estruturas avaliadas apenas *in silico*, enquanto que as estruturas estudadas experimentalmente são 2o5u, 2av9 são estruturas de tioesterase descritas para *Pseudomonas aeruginosa*, e são relatadas no mesmo artigo científico (CHRUSZCZ et al.2008) e a terceira é estrutura descrita para *T. thermophilus* 2cye (HOSAKA et al.2009).

Atualmente existem vários estudos em busca do melhor entendimento da função biológica dessas proteínas, pois algumas são patogênicas humanas e por se tratar de bactérias gram-negativas, o entendimento dessas proteínas pode ajudar no desenvolvimento de novos medicamentos.

Tabela 8. Predição dos modelos para TEs-Meta

Código PDB	Função	Cobertura	ID %	Método	Formação estrutural	Ligantes *
2o5u <i>P. aeruginosa</i>	Tioesterase	0,72	23,74	X-ray, 1.9 Å	homo tetramero	—
2nuj <i>Jannaschia sp.</i>	Tioesterase	0,74	21,13	X-ray 2 Å	homo tetramero	-
2av9 <i>P. aeruginosa</i>	Tioesterase	0,72	23,91	X-ray 2.4 Å	homo tetramero	—
3hmo <i>Bartonella henselae</i>	Provável Tioesterase	0,70	22,22	X-ray 2.5 Å	homo tetramero	—
2cye <i>T. thermophilus</i>	Putative Tioesterase	0,67	33,59	X-ray 1.9 Å	homo tetramero	CoA, Zn
1s5u <i>E. coli</i>	Proteína ybgC	0,69	22,63	X-ray 1.7 Å	homo tetramero	—
1z54 <i>T. thermophilus</i>	Provável Tioesterase	0,67	28,91	X-ray 2.1 Å	homo tetramero	-
2egi <i>Aquifex aeolicus</i>	Putative Tioesterase	0,65	30,40	X-ray 2.3 Å	homo tetramero	—
2gt6 <i>Sulfolobus solfataricus</i>	Proteína hipotética	0,67	23,26	X-ray 1.9 Å	homo tetramero	CoA
2oiw <i>B. stearothermophilus</i>	putative 4-hidroxibenzoil-CoA tioesterase	0,67	22,66		homo tetramero	Mg
4k00 <i>Synechocystis</i>	1,4-dihidroxi-2-naphthoil-CoA tioesterase	0,68	18,32	X-ray 1.9 Å	homo tetramero	—

*CoA acetil coenzima A, Zn íon zinco, Mg íon magnésio.

Após a predição estrutural da TEs-Meta (Figura 23), foram realizados testes de confiabilidade do modelo através do gráfico de *Ramachandran*

A figura 25 possui características equivalentes às descritas anteriormente para família Hotdog-fold, que possui a estrutura formada por uma hélice central, envolto por seis de cadeia antiparalelas de folhas-beta estrutura já bem definida pela literatura (DILLON e BATEMAN, 2002, ANGELINI et al 2008, PIDUGU et al.,2009).

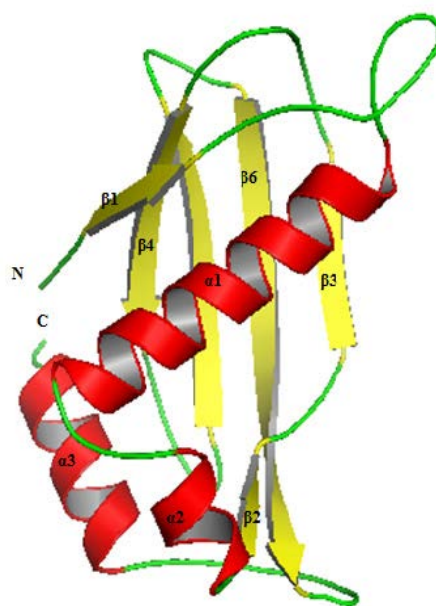


Figura 25. Predição estrutural TEs-Meta: As estruturas em vermelho representam as α -hélices, em amarelo as folhas- β e em verde o loop. A numeração das folhas- β de 1 a 6 seguem a orientação no sentido N para o C terminal. Estrutura desenhada no programa Pymol version 1.5.0.4.

O gráfico gerado para TEs-Meta apresentou 93,2% aminoácidos localizados na região mais favorável, representada pela região azul escura do gráfico de *Ramachandran*, e a maior parte dos resíduos localizados na região do gráfico que representa as estruturas secundárias equivalente às folhas- β antiparalelas, possui 4,3% de resíduos na região favorável glicina, de uma região flexível para a conformação dos resíduos e apenas 2,3% ficou na região não permitida, de acordo com o gráfico *Ramachandran*, a Tes-Meta possui confiabilidade estrutural, onde requer pouco gasto energético para manter sua conformação (Figura 26).

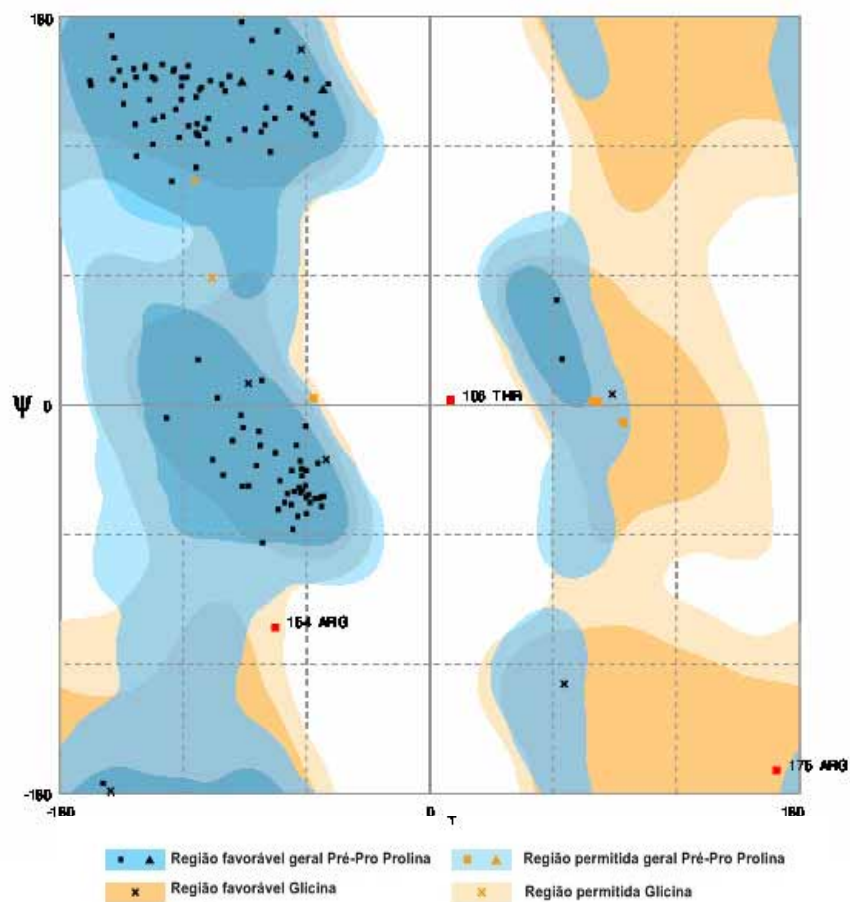


Figura 26. Gráfico Ramachandran baseado nos ângulos de torção ϕ e ψ para TEs- Meta: : a região azul escura representa região favorável geral Pré-Pró/Prolina a azul clara região permitida geral Pré-Pró/Prolina; em laranja escuro região favorável geral para Glicina e laranja claro região permitida geral para Glicina. Gráfico foi gerado no programa Rampage.

A figura 27 foi gerada no ProSA web (WIEDERSTEIN, SIPPL, 2007) que indica o valor de z-score avaliando a para TEs-Meta $-4,28$ com 135 aminoácidos representa uma de massa molecular pequena sugerindo que o valor de z-score negativo encontrado pode ser atribuído a publicação de estruturas erradas em publicadas em bancos de dados e, com a sequências incorretas (WIEDERSTEIN e SIPPL, 2007; BENKERT et al., 2011).

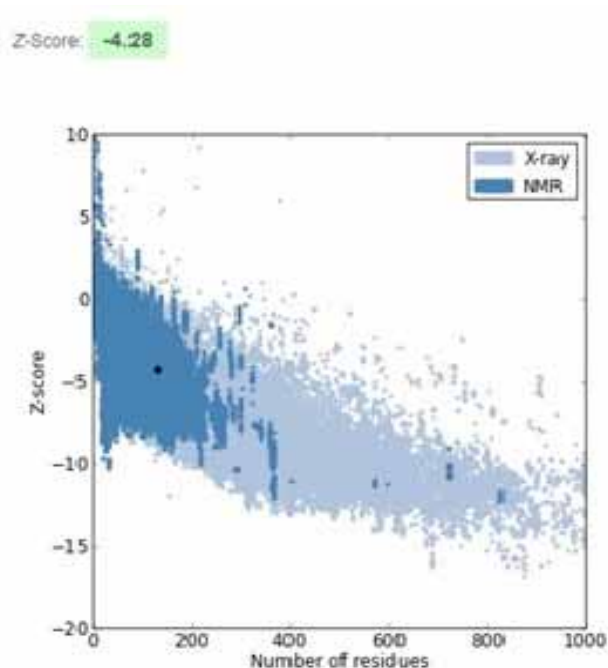


Figura 27. Pontuação de Z-score para TEs-Meta: pontos indicados com azul escuro são estruturas de proteína resolvidas por ressonância magnética nuclear (NMR) e pontos em azul claro corresponde as estruturas resolvidas através cristalografia de raios-X, o círculo preto mostra no gráfico o local da pontuação para TEs-Meta. (ProSa Web).

Após a avaliação estrutural, a sequência molde e modelo foram sobrepostas e o valor de rmsd 1.346 Å, pôde assumir que a TEs-Meta possui estrutura bem próxima ao molde (Figura 28 A). As figuras 28 B e 28 C indicam o possível sítio de ligação para TEs-Meta onde esta localizada a estrutura da coenzima A, além disso, é nesse ponto que outros monômeros de tioesterase podem se unir para formação da estrutura quaternária. A figura 28 C ressalta a posição dos ligantes Thr116, Ile117, Glu118, Gly143, Ala144, Arg145 e Pro176.

Segundo OHTSUKA (2009) esses resíduos estão presentes no modelo cristalográfico do *T. thermophilus*, a presença do ácido glutâmico (Glu118) pode ser indicio que TEs-Meta pertence a família Hotdog-fold e a subfamília Acyl-CoA tioester hidrolase que é dividida em duas classes, *YbgC* que pertence ao cluster *tol-pal* que é encontrado no sistema patogênico de bactérias gram-negativas que, e possui papel importante na manutenção da integridade do envelope celular e, provavelmente esta envolvida no metabolismo de fosfolipídio (ANGELINI et al., 2008; COHEN, 2013) a e *ybaW* faz parte do processo de regulação β - oxidação (FENG e CRONAN, 2012).

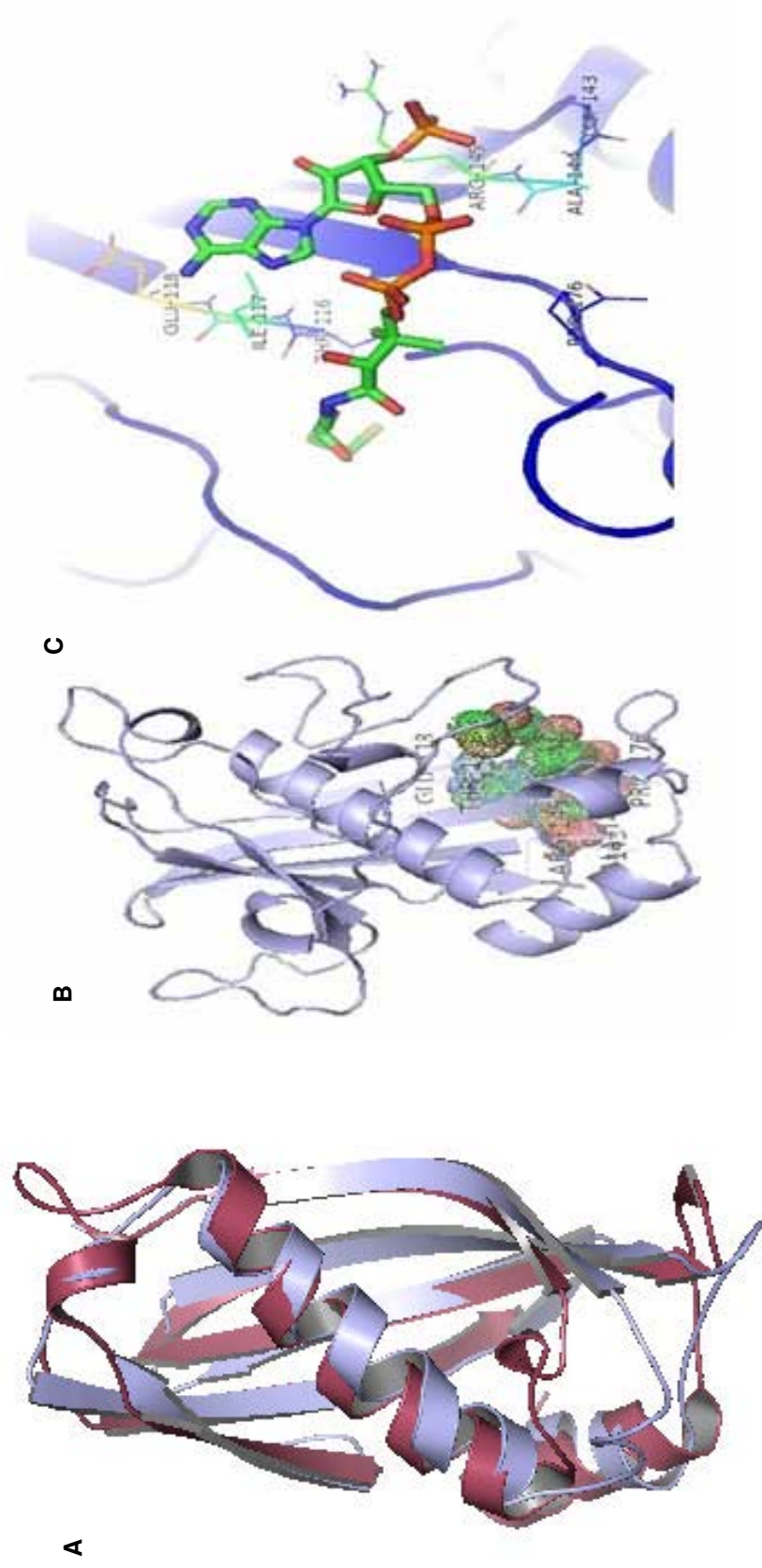


Figura 28. Predição estrutural TEs-Meta: A) sobreposição de TEs-Meta na cor lilás e *T. thermophilus* vinho. B) Visão geral da estrutura TEs-Meta. C) Destaque para sítios de ligação, estrutura no centro representa coenzima A, resíduos ligados a CoA estão marcados com número da posição de ligação (Thr116, Ile117, Glu118, Gly143, Ala144, Arg145 e Pro176). Estruturas desenhadas no programa Pymol version 1.5.0.4.

Sabendo que TEs-Meta é um Hotdog-fold é importante estabelecer a qual subfamília TEs-Meta ela pertence, nesse sentido foi utilizada a ferramenta PSIPRED, descrito no item 5.2.1.

O motivo conservado para TEs-Meta [RVY-X₂-DTD-X₂-GVV-X-H-X₂Y], e o similar é da subfamília YbgC [DTD-X₂-GVV-X-H-X₂Y] (DUBUISSON et al., 2005, ANGELINI et al., 2008; PIDUGU et al., 2009). A subfamília YbgC já foi descrita anteriormente, mas além dos motivos conservados serem similares a TEs-Meta possui o resíduo ácido glutâmico como próximo ao sitio ativo.

Após identificação do motivo conservado foi realizado alinhamento entre a TEs-Meta e alguns membros representantes da subfamília YbgC, para identificação no alinhamento dos pontos descritos como possível domínio conservado (Figura 29).

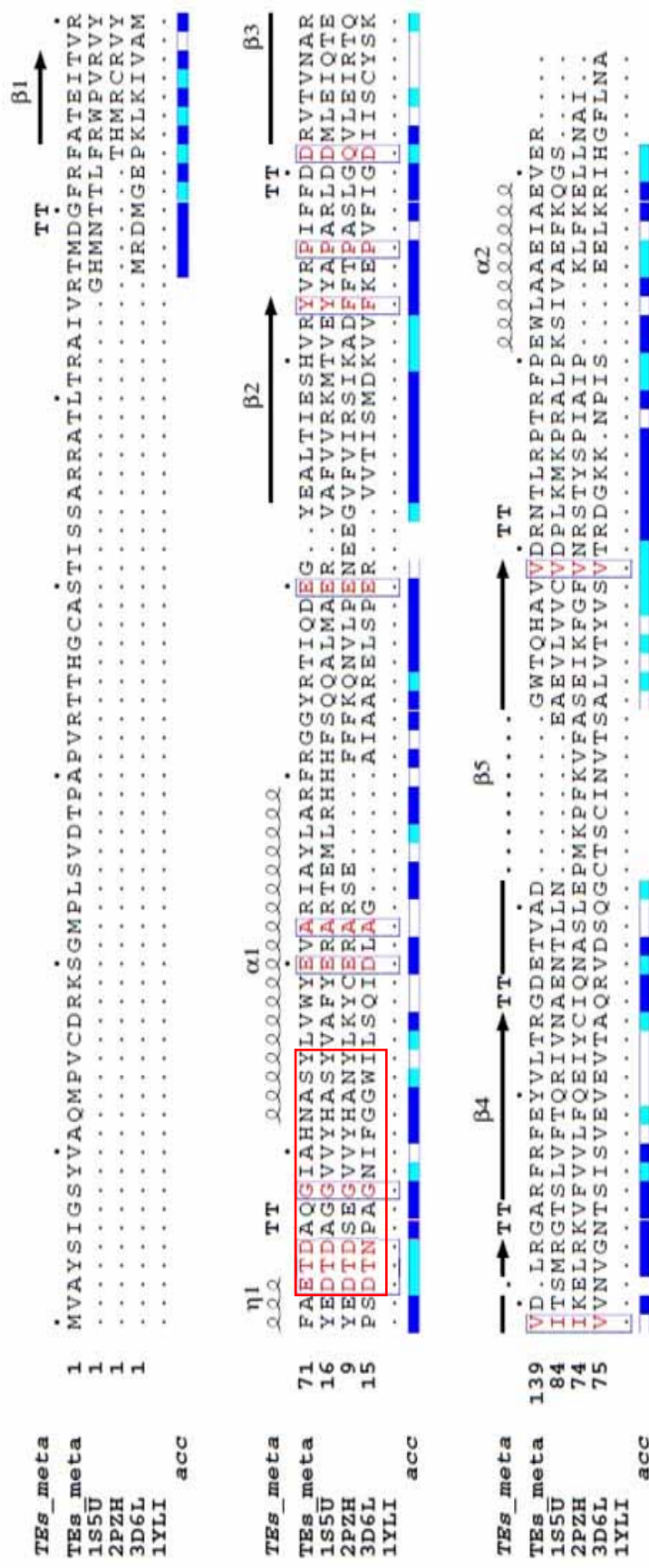


Figura 29. Alinhamento múltiplo de sequências TES-Meta : resíduos conservados estão marcados em vermelho, e regiões com resíduos homólogos estão em marcados na caixa em azul, o quadro vermelho marca as sequências relacionadas ao motivo da família YbgC. Acima as regiões irregulares representam em espiral as α -hélices, e as setas folhas beta. As sequências de aminoácidos usadas na análise são: 1S5U – *E. coli*, 2PZH *Helicobacter pylori*, 3D6L *Cryptosporidium hominis*, 1YLI *Haemophilus influenzae*.

O alinhamento indica que existe uma variação evolutiva e, mostraram ainda que os resíduos delimitados pelos quadros azuis, também possuem substituição de aminoácidos com caráter conservativo, substituição esta que ocorreu entre resíduos com propriedades químicas semelhantes (Glu por Asp).

Após a avaliação do alinhamento foi construído o fenograma, onde foram usadas sequências de aminoácidos representativas da tioesterase selecionadas no NCBI (Figura 30).

TEs-Meta foi agrupada entre os tetrâmeros (TA), e possui similaridade como o gene 1Z54 (*T. thermophilus*), por estar no mesmo clado, 1Z54, pertencente aos microrganismos da subfamília YbgC. (PIDUGU et al. 2009).

Segundo Pidugu (2009) a análise revela características distintas, o grupo tetrâmero (TA) aparece em diferentes subfamílias: YbgC, YbaW e 4HBT-I, o tipo TB de tetrâmeros são encontrados nas subfamílias TEs B-like tioesterase, 4HBT-II.

As análises fenética revelou que a TEs-Meta representa um membro da subfamília YbgC. Esta suposição é baseada no fato de que os dados obtidos corroboram com os da literatura (DILLON e BATEMAN, 2004, DUBUISSON et al., 2005, ANGELINI et al., 2008, OHTSUKA et al., 2009, PIDUGU et al., 2009).

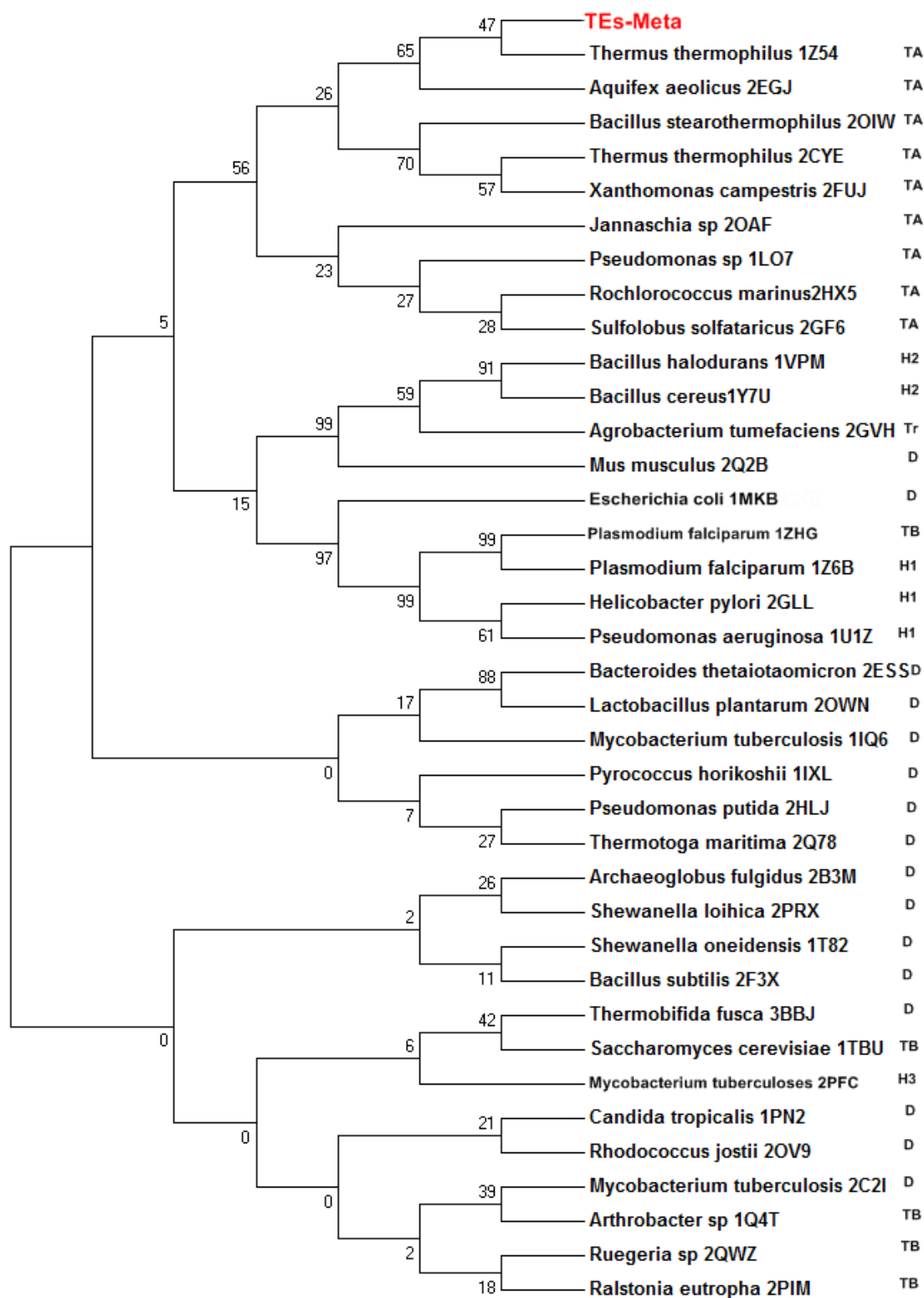


Figura 30. Análise fenética de sequências proteicas hot-dog fold: As sequências foram selecionadas de acordo com similaridade genes de proteínas “hotdog fold.” As siglas D- dímeros; TA- tetrâmeros, TB-tetrâmeros com sobreposição de folhas β , H1- hexâmero com sítio ativo na interface do loop; H2 –hexâmero com hélice na interface N-Terminal; H3- hexâmero com pareamento entre dímeros e Trimêros.

É importante salientar ainda que os membros dessas famílias e subfamílias ainda não foram totalmente caracterizados prejudicando assim avaliação. A falta de informação sobre as estruturas pode causar redundância entre os resultados, pois a maioria dos microrganismos estudados e os dados publicados aparecem no banco com diferentes números de acesso.

5.3.1 Clonagem em vetor de expressão

As sequências codificadoras da pirofosfatase e tioesterase identificadas nas etapas anteriores foram utilizadas como molde para a construção de iniciadores visando à amplificação desses genes para clonagem em vetor de expressão de proteínas. Foram construídos iniciadores para os genes codificadores das proteínas MetaPPIase e TEs-Meta (Tabela 2). A amplificação foi bem sucedida apenas no caso da TEs-Meta gerando fragmentos de aproximadamente 583 pb. O produto de PCR foi purificado, digerido com as respectivas endonuclease (sítios presentes nos iniciadores) e clonados no vetor de expressão pET28b digerido com as mesmas endonucleases. A identidade do gene clonado foi confirmada por perfil de restrição (Figura 31 A,B) e sequenciamento (Figura 32).

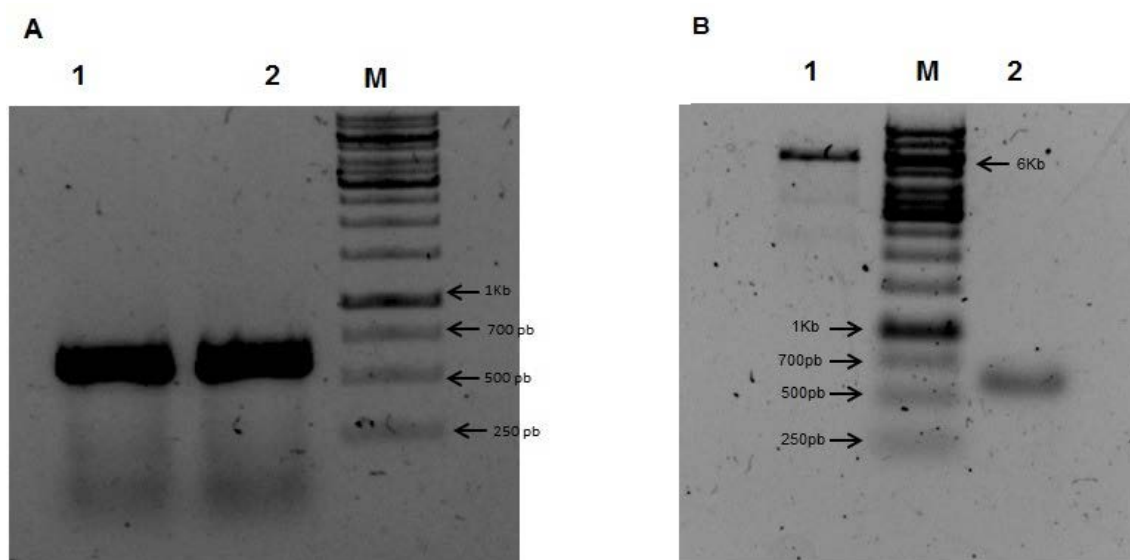


Figura 31. Eletroforese em gel de agarose 1% dos produtos de PCR e Perfil de restrição dos plasmídeos recombinantes em vetor pET28b contendo os genes codificadores de TEs-Meta :: A: as amostras 1 e 2 são amplicons do gene TEs-Meta em duplicata e M

marcador tamanho de Molecular 1 kb DNA “ladder” (Fermentas). B) 1 o é o vetor pET28a e 2 corresponde ao TEs-Meta.

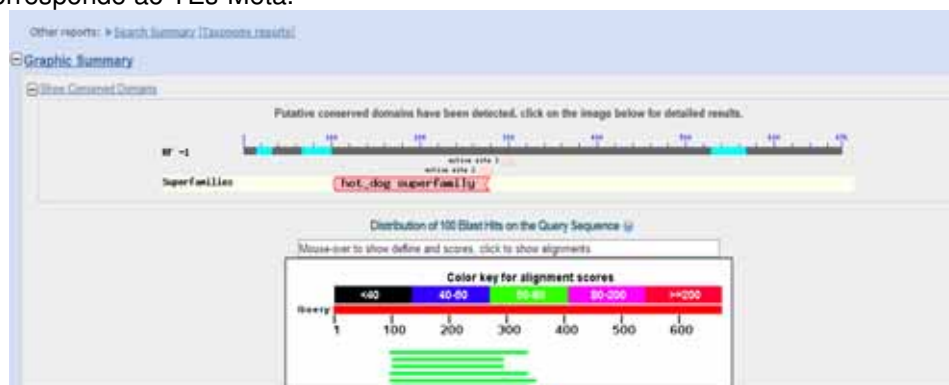


Figura 32. Confirmação da clonagem do inserto. A sequência do gene clonado foi analisado no BlastX.

5.3.2 Expressão e Purificação

A expressão da proteína foi avaliada no extrato celular bruto em gel desnaturante SDS antes e após indução com IPTG. Foi possível visualizar uma banda de proteína expressa após a indução durante 22h (Figura 32) indicando que o sistema de expressão, formado pela célula hospedeira, vetor pET28b e o gene clonado. O tamanho aproximado da proteína expressa é de aproximadamente 25 kDa (Figura 18).

A proteína TEs-Meta foi purificada por cromatografia de afinidade em coluna de níquel, por meio da interação entre os resíduos de histidina (His-tag) adicionados na região N-terminal da proteína e os íons metálicos de níquel da coluna. A proteína foi eluída em tampão indicado pelo fabricante com concentrações crescentes de imidazol com o pico de eluição em 500 mM. Foram coletadas 3 frações de 5 ml, e a fração proteica obtida foi concentrada para 500µL (volume final) usando filtro termo (Figura 32).

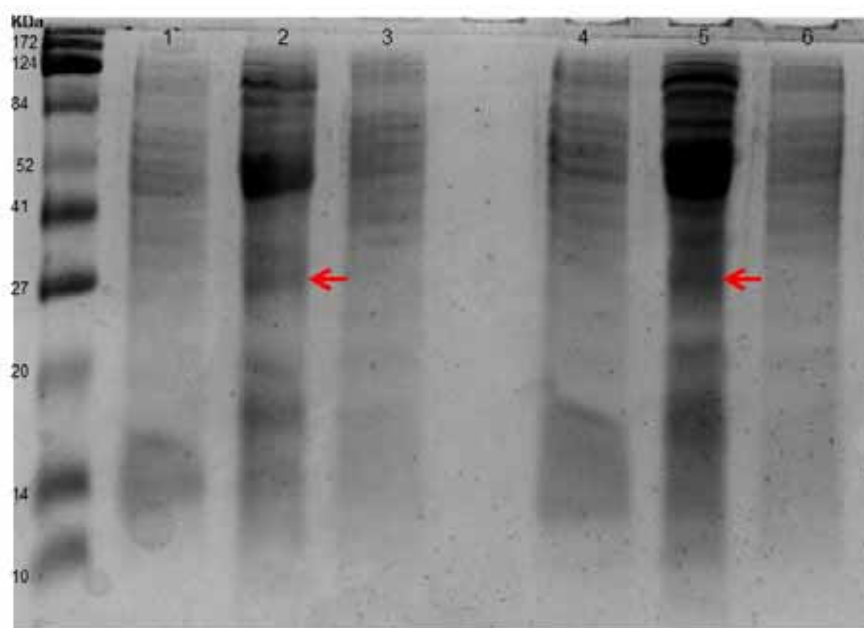


Figura 33. Eletroforese em SDS-PAGE a 12% dos extratos celulares de *E. coli* expressando TEs-Meta: M – marcador molecular Lonza (127 kDa to 10 kDa), as amostras 1 e 4 TEs wash 1 (lavagem) sem imidazol, amostra 2-proteína filtrada 500 mM imidazol e de 3 – 6 Proteína total e 5 – proteína filtrada e concentrada.

A Figura 33 demonstra a presença da proteína TEs-Meta de massa molecular de aproximadamente 25 KDa, presente apenas na amostra 2 e 5 mas, ainda não está em condição de ideais de expressão, assim é importante fazer adaptações da expressão ou mudança do inserto para outro vetor a fim de otimizar a expressão da proteína TEs-Meta para dar continuidade aos estudos de caracterização estrutural e cinética.

6. CONCLUSÕES

Nossos resultados confirmam a hipótese de que a elevada diversidade microbiana do solo permite acessar novos genes por meio da abordagem metagenômica.

Anotação funcional dos genes permitiu a identificação de dois diferentes genes que codificam as enzimas MetaPPiase e a TEs-Meta, que apresentam funções inéditas ainda não descritas na literatura.

A MetaPPiase foi classificada como sendo uma proteína integral de membrana capaz de atuar como H⁺ PPiase pertencente à subfamília de K⁺ independente. Para a confirmação da predição é necessária clonagem no vetor de expressão posterior caracterização cinética e cristalização.

A TEs-Meta foi classificada como membro tioesterase tipo II pertencente ao grupo *hot-dog fold* e associada a subfamília YbgC. Os ensaios de expressão proteica realizados mostram que ainda são necessárias adaptações na metodologia para conseguir expressão que permita sua caracterização estrutural e cinética.

As ferramentas de bioinformática foram essenciais para o entendimento e classificação dos genes selecionados.

7- REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER P. Molecular biology of the cell, fourth ed., **Garland Science**, New York, 2002.

ALPHEY, L. DNA Sequencing: From Experimental Methods to Bioinformatics. New York: Springer, 1997. Capítulo: 17: **Protein Structure Prediction**, p. 179. ISBN 0-387-91509-5

ANGELINI, A.; CENDRON, L.; GONCALVES, S.; ZANOTTI, G.; TERRADO, L. Structural and enzymatic characterization of HP0496, a YbgC thioesterase from *Helicobacter pylori*. **Proteins: Structure, Function, and Bioinformatics**, 72(4), 1212-1221, 2008.

ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. **Bioinformatics**, 22:195-201, 2006.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, 25(17), 3389-3402, 1997.

BAYKOV, A. A.; MALINEN, A. M.; LUOTO, H.H.; LAHTI, R. Pyrophosphate-fueled Na⁺ and H⁺ transport in prokaryotes. **Microbiology and Molecular Biology Reviews**, 77, 267-276, 2013.
<http://dx.doi.org/10.1128/MMBR.00003-13>

BENKERT, P.; BIASINI, M.; SCHWEDE, T. Toward the estimation of the absolute quality of individual protein structure models. **Bioinformatics**, 27(3), 343-350, 2011.

BENNER, S. A.; CANNAROZZI, G.; GERLOFF, D.; TURCOTTE, M.; CHELVANAYAGAM, G. Bona fide predictions of protein secondary structure using transparent analyses of multiple sequence alignments. **Chemical Reviews**, 97(8), 2725-2844, 1997.

BINOD, P.; PALKHIWALA, P.; GAIKAIWARI, R.; RAGHAVENDRA, N.; NAMPOOTHIRI, K.; M.; DUGGAL, A.; DEY, K.; PANDEY, A. Industrial Enzymes-Present status and future perspectives for India. **Journal of Scientific & Industrial Research**, 72:271-286, 2013.

BORDOLI, L.; KIEFER, F.; ARNOLD, K.; BENKERT, P.; BATTEY, J.; SCHWEDE, T. Protein structure homology modelling using SWISS-MODEL Workspace. **Nature Protocols**, 4,1, 2009.

BOWIE, J. U.; LUTHY, R.; EISENBERG, D. A method to identify protein sequences that fold into a known three-dimensional structure. **Science**, 253 (5016), 164-170, 1991.

CANTU, D. C.; CHEN, Y.; REILLY, P. J. Thioesterases: a new perspective based on their primary and tertiary structures. **Protein Science**, v. 19(7), 1281-1295, 2010.

CARY, S.C.; MCDONALD, I.R.; BARRETT, J.E.; COWAN, D.A. On the rocks: the microbiology of Antarctic Dry Valley soils. **Nat Rev Microbiol** 8: 129-138, 2010.

CHRUSZCZ, M.; ZIMMERMAN, M. D.; WANG, S.; KOCLEGA, K. D.; ZHENG, H.; EVDOKIMOVA, E.; MINOR, W. Function-biased choice of additives for optimization of protein crystallization: the case of the putative thioesterase PA5185 from *Pseudomonas aeruginosa* PAO1. **Crystal Growth and Design**, 8(11), 4054-4061, 2008.

COHEN, D. E. New players on the metabolic stage. **Adipocyte** 2(1): 3-6, 2013.<doi:10.4161/adip.21853>

CORRÊA, A. G.; ZUIN, V. G., Química Verde: Fundamentos e Aplicações. 1 ed. São Carlos: **EDUFSCar**, v. 2, p107-110, 2012.

DICK, G. J.; TEBO, B. M. Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. **Environ Microbiol**. 2010.

DILLON, S. C.; BATEMAN, A. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. **BMC bioinformatics**, 5:109, 2004.

DUBUISSON, J. F; VIANNEY, A; HUGOUVIEUX-COTTE-PATTAT, N; LAZZARONI, J. Tol-Pal proteins are critical cell envelope components of *Erwinia chrysanthemi* affecting cell morphology and virulence. **Microbiology**, 151:3337-3347, 2005.

FABRICHNIY, I. P.; LEHTIÖ, L.; TAMMENKOSKI, M.; ZYRYANOV, A. B.; OKSANEN, E.; BAYKOV, A. A.; LAHTI, R. GOLDMAN, A. A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. **Journal of Biological Chemistry**, 282(2), 1422, 2007.

FELIX, H.; FLEISCH, H. Properties of inorganic pyrophosphatase of pig scapula

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, 783-791, 1985.

FENG, Y.; CRONAN, J. E. A new member of the *Escherichia coli* fad regulon: transcriptional regulation of fadM (ybaW). **Journal of bacteriology**, 191(20), 6320-6328, 2009.

FENG, Y.; CRONAN, J. E. Crosstalk of *Escherichia coli* FadR with global regulators in expression of fatty acid transport genes. **PloS one**, 7:e46275, 2012.

FERRER, M.; MARTÍNEZ-ABARCA, F.; GOLYSHIN, P. N. Mining genomes and 'metagenomes' for novel catalysts. **Current Opinion in Biotechnology**, 16(6), 588-593, 2005.

GOUET, P.; COURCELLE, E.; STUART, D. I. ESPript: analysis of multiple sequence alignments in PostScript. **Bioinformatics**, 15:305-308, 1999. <http://esprpt.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>.

GUEx, N.; PEITSCH, M.C.; SCHWEDE, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb Viewer: a historical perspective. **Electrophoresis** 30(Suppl 1):S162–73, 2009.

HALL, P. BioEdit –version 5.0.6. Raleigh, North Carolina State University, Department of Microbiology, 2001.

HANS-DIETER, H.; FOLKERS, G. Molecular modeling: basic principles and applications. **John Wiley & Sons**, 2008.

HANDELSMAN, J. Metagenomics: Application of genomics to uncultured microorganisms. **Microbiology and Molecular Biology Reviews**, v 68, p. 669-685, 2004.

HANDELSMAN, J.; RONDON, M. R.; BRADY, S. F.; CLARDY, J.; GOODMAN, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chemistry and Biology**, v. 5, p. 245-249, 1998.

HUNTER-CEVERA. The value of microbial diversity. **Curr Opin Microbiol.** 1:278-285, 1998.

HOSAKA, T.; MURAYAMA, K.; KATO-MURAYAMA, M.; URUSHIBATA, A.; AKASAKA, R.; TERADA, T.; YOKOYAMA, S. Structure of the putative thioesterase protein TTHA1846 from *Thermus thermophilus* HB8 complexed with coenzyme A and a zinc ion. **Acta Crystallographica Section D: Biological Crystallography**, 65(8), 767-776, 2009.

IQBAL, H. A.; FENG, Z.; BRADY, S.F. Biocatalysts and small molecule products from metagenomic studies. **Current Opinion in Chemical Biology**, 16 (1-2), pp. 109-116, 2012.

JÄMSEN, J.; BAYKOV, A. A.; LAHTI, R. Fast kinetics of nucleotide binding to *Clostridium perfringens* Family II Pyrophosphatase containing CBS and DRTGG domains. *Biochemistry (Moscow)*, vol. 77, p. 205-211, 2012.

KAJANDER, T.; KELLOSALO, J.; GOLDMAN, A. Inorganic pyrophosphatases: One substrate, three mechanisms. **FEBS Letters**, 587, 1863-1869, 2013. <<http://dx.doi.org/10.1016/j.febslet.2013.05.003>>

KELLOSALO, J.; KAJANDER, T.; KOGAN, K.; POKHAREL, K.; GOLDMAN, A. The structure and catalytic cycle of a sodium pumping pyrophosphatase. **Science**, 337, 473-476, 2012. <<http://dx.doi.org/10.1126/science.1222505>>

KURILOVA, S. A.; BOGDANOVA, A. V.; NAZAROVA, T. I.; AVAEVA, S. M. Changes in the *E. coli* inorganic pyrophosphatase activity on interaction with magnesium, zinc, calcium and fluoride ions. **Bioorg. Khim.** 10:1153-1160, 1984.

LABONTE, J. W.; TOWNSEND, C. A. Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. **ACS Publication**, 2182-2204, 2012.

LATHAM, J. A. **Structure to function: case studies of hotdog-fold superfamily thioesterases from *Escherichia coli***, 254 p. Dissertation (PhD in Chemistry) – University of New Mexico, Albuquerque, New Mexico, 2012.

LEE, M. H.; LEE, S. W. Bioprospecting Potential of the Soil Metagenome: Novel Enzymes and Bioactivities. **Genomics & informatics**, 11(3), 114-120, 2013.

LI, S.; YANG, X.; YANG, S.; ZHU, M.; WANG, X. Technology prospecting on enzymes: Application, marketing and engineering. **Computational and Structural Biotechnology Journal**, 2, 2012.

LIN, S.M.; TSAI, J.Y.; HSIAO, C.D.; HUANG, Y.T.; CHIU, C.L.; LIU, M.H.; TUNG, J.Y.; LIU, T.H.; PAN, R.L.; SUN, Y.J. Crystal structure of membrane-embedded H⁺-translocating pyrophosphatase. **Nature**, 484, 399-403, 2012. <http://dx.doi.org/10.1038/nature10963>

LUOTO, H. H.; BELOGUROV, G. A.; BAYKOV, A. A.; LAHTI, R.; MALINEN, A. M. Na⁺-translocating membrane pyrophosphatases are widespread in the microbial world and evolutionarily precede H⁺-translocating pyrophosphatases. **Journal of Biological Chemistry**, 286, 21633-21642, 2011. <http://dx.doi.org/10.1074/jbc.M111.244483>

MENDE, D. R.; WALLER, A. S.; SUNAGAWA, S.; JÄRVELIN, A. I.; CHAN, M. M.; ARUMUGAM, M.; BORK, P. Assessment of metagenomic assembly using simulated next generation sequencing data. **PloS one**, 7(2), e31386, 2012.

MIMURA, H.; NAKANISHI, Y.; HIRONO, M.; MAESHIMA, M. Membrane topology of the H⁺-pyrophosphatase of *Streptomyces coelicolor* determined by cysteine-scanning mutagenesis. **The Journal of Biological Chemistry**, 279,

MITCHELL, S. J.; MINNICK, M. F. Cloning, functional expression, and complementation analysis of an inorganic pyrophosphatase from *Bartonella bacilliformis*. **Can. J. Microbiol.** 43:734-743, 1997.

MORRIS, A. L.; MACARTHUR, M. W.; HUTCHINSON, E.G.; THORNTON, J. M. Stereochemical quality of protein structure coordinates. **Proteins**, 12:345-364, 1992.

MOUNT, D. W. Bioinformatics: Sequence and genome analysis, second ed., **Cold Spring Harbor Press**, Cold Spring Harbor, NY, 2004.

NIE, L.; REN, Y.; SCHULZ, H. Identification and Characterization of *Escherichia coli* Thioesterase III That Functions in Fatty Acid β -Oxidation. **Biochemistry**, 47(29), 7744-7751, 2008 a.

NIE, L.; REN, Y.; JANAKIRAMAN, A.; SMITH, S.; SCHULZ, H. A novel paradigm of fatty acid beta-oxidation exemplified by the thioesterase-dependent partial degradation of conjugated linoleic acid that fully supports growth of *Escherichia coli*, **Biochemistry**, 47, 9618-26, 2008 b.

NELSON, D.L.; COX, M.M. Princípios de bioquímica de Lehninger. 5.^a ed. Porto Alegre, Editora Artmed, 2011.

NOWROUSIAN, M. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. **Eukaryotic cell**, v. 9, n. 9, p. 1300-1310, 2010.

NUGENT, T.; WARD, S.; JONES, D. T. The MEMPACK alpha-helical transmembrane protein structure prediction server. **Bioinformatics**, 27, 1438-1439, 2011. <http://dx.doi.org/10.1093/bioinformatics/btr096>

OHTSUKA, J.; ICHIHARA, Y.; EBIHARA, A.; NAGATA, K.; TANOKURA, M. Crystal structure of TTHA1264, a putative M16-family zinc peptidase from *Thermus thermophilus* HB8 that is homologous to the β subunit of mitochondrial processing peptidase. **Proteins: Structure, Function, and Bioinformatics**, 75(3), 774-780, 2009.

OLIVEIRA, V. M.; SETTE, L. D.; GARBOGGINI, F. F. Preservação e Prospecção de Recursos Microbianos. **Multiciência**, v. 7, p. 1-19, 2006.

OKSANEN, E.; AHONEN, A. K.; TUOMINEN, H.; TUOMINEN, V.; LAHTI, R.; GOLDMAN, A.; HEIKINHEIMO, P. A complete structural description of the catalytic cycle of yeast pyrophosphatase. **Biochemistry** 46:1228-1239, 2007.

PACE, N. R. A molecular view of microbial diversity and the biosphere. **Science**. 276:734-740, 1997.

PARFENYEV, A. N.; SALMINEN, A.; HALONEN, P.; HACHIMORI, A.; BAYKOV, A. A.; LAHTI, R. Quaternary structure and metal ion requirement of family II pyrophosphatases from *Bacillus subtilis*, *Streptococcus gordonii*, and *Streptococcus mutans*. **J. Mol. Biol.** 276:24511-24518, 2001.

PEARCE, D. A.; HUGHES, K. A.; LACHLAN-COPE, T.; HARANGOZO, S. A.; JONES, A. E. Biodiversity of air-borne microorganisms at Halley station, **Antarctica. Extremophiles** 14: 145-159, 2010.

PIDUGU, L. S.; MAITY, K.; RAMASWAMY, K.; SUROLIA, N.; SUGUNA, K. Analysis of proteins with the 'hot dog' fold: Prediction of function and identification of catalytic residues of hypothetical proteins. **BMC structural biology**, 9(1), 37, 2009.

PRAKASH, T.; TAYLOR, T. D. Functional assignment of metagenomic data: challenges and applications. **Briefings in bioinformatics**, 13(6), 711-727, 2012.

PUNTA, M.; COGGILL, P.C.; EBERHARDT, R.Y.; MISTRY, J.; TATE, J.; BOURSNEILL, C.; PANG, N.; FORSLUND, K.; CERIC, G.; CLEMENTS, J.; HEGER, A.; HOLM, L.; SONNHAMMER, E.L.L.; EDDY, S.R.; BATEMAN, S.; FINN, R.D. The Pfam protein families database. **Nucleic acids research**, 40(D1), D290-D301, 2012.

RAAIJMAKERS, J. M.; VLAMI, M.; DE SOUZA, J. T. Antibiotic production by bacterial biocontrol agents. **Antonie van Leeuwenhoek**, 81(1-4), 537-54, 2002.

REN, Y.; AGUIRRE, J.; NTAMACK, A. G.; CHU, C.; SCHULZ, H. An alternative pathway of oleate beta-oxidation in *Escherichia coli* involving the hydrolysis of a dead end intermediate by a thioesterase, **The Journal of Biological Chemistry** 279, 11042-50, 2004.

RODRIGUES, G.R., **Prospecção de genes de celulase presentes em biblioteca metagenômica**, 47p. Dissertação (Mestrado em Microbiologia Agropecuária) Universidade Estadual Paulista UNESP-FCAV, Jaboticabal, 2008.

RODRIGUES, G. R.; VAL-MORAES, S. P.; DE MACEDO LEMOS, E. G.; PIZAULO, J. M. Novel Inorganic Pyrophosphatase from Soil Metagenomic and Family and Subfamily Prediction. **Open Journal of Applied Sciences**, 2014.

< <http://dx.doi.org/10.4236/ojapps.2014.42008> >

RODRÍGUEZ-GUILBE, M.; OYOLA-ROBLES, D.; SCHREITER, E. R.; BAERGA-ORTIZ, A. Structure, Activity, and Substrate Selectivity of the Orf6 Thioesterase from *Photobacterium profundum*. **Journal of Biological Chemistry**, 288(15), 10841-10848, 2013.

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

SAHA, P.; KRISHNAMURTHI, S.; BHATTACHARYA, A.; SHARMA, R.; CHAKRABARTI, T. *Fontibacillus aquaticus* gen. nov., sp. nov., isolated from a warm spring. **Int J Syst Evol Microbiol** 60: 422-428, 2010.

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

SAMBROOK, J; RUSSELL, D.W; RUSSELL, D.W. MOLECULAR CLONING: A laboratory manual. **Cold Spring Harbour Laboratory Press**, New York. 2001.

SAMYGINA, V. R.; POPOV, A. N.; RODINA, E. V.; VOROBYEVA, N. N.; LAMZIN, V. S.; POLYAKOV, K. M.; KURILOVA, S. A.; NAZAROVA, T. I.; AVAEVA, S. M. The structures of *E. coli* inorganic pyrophosphatase complexed with Ca^{2+} or Ca PPI at atomic resolution and their mechanistic implications. **J. Mol. Biol.** 314:633- 645, 2001.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. de. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253-259, 2003.

SARROUH, B.; MARCOS SANTOS, T.; MIYOSHI, A.; DIAS, R.; AZEVEDO, V. Up-to-date insight on industrial enzymes applications and global market. **Journal of Bioprocessing & Biotechniques** S, 4, 2012.

SERVANT, F.; ABAJIAN, F.; GOUZY, J. KAHN, D. ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. **Nucleic acids research**, 28(1), 267-269, 2000.

SIMONS, K. T.; KOOPERBERG, C.; HUANG, E.; BAKER, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. **Journal of molecular biology**, 268(1), 209-225, 1997.

SCHECKENBACH, F.; HAUSMANN, K.; WYLEZICH, C.; WEITERE, M.; ARNDT, H. Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. **Proc Natl Acad Sci U S A** 107: 115-120, 2010.

SCHRÖDINGER, L.L.C. The PyMOL Molecular Graphics System, Version1, 2010.

SCHUCH, V.; GOMES, E. S.; LEMOS, E. G. M. Discovery of genes related to antibiotic biosynthesis into a metagenomic library. **New Biotechnology**25:S105, 2009.

SHINTANI, T.; UCHIUMI, T.; YONEZAWA, T.; SALMINEN, A.; BAYKOV, A. A.; LAHTI, R.; HACHIMORI, A. Cloning and expression of a unique inorganic pyrophosphatase from *Bacillus subtilis* :evidence for a new family of enzymes. **FEBS letters**, 439(3), 263-266, 1998.

SILVEIRA, É. L. D.; PEREIRA, R. M.; SCAQUITTO, D.C.; PEDRINHO, E. A. N.; VAL-MORAES, S. P.; WICKERT, E; CARARETO-ALVES, L. M.; LEMOS, E. G. D. M. Bacterial diversity of soil under eucalyptus assessed by 16S rDNA sequencing analysis. **Pesquisa Agropecuária Brasileira** 41:1507-1516, 2006. <<http://dx.doi.org/10.1590/S0100-204X2006001000008>>

SIMON, C.; DANIEL, R. Achievements and new knowledge unraveled by metagenomic approaches. **Applied microbiology and biotechnology**, 85(2), 265-276, 2009.

SIGRIST, C.J.A.; DE CASTRO E.; CERUTTI, L.; CUCHE, B.A.; HULO, N.; BRIDGE, A.; BOUGUELERET, L.; XENARIOS, I. New and continuing developments at PROSITE, **Nucleic Acids Res.** 2012; <doi: 10.1093/nar/gks1067>

SONG, Z. Q.; CHEN, J. Q.; JIANG, H. C.; ZHOU, E. M.; TANG, S. K.; ZHI, X. Y.; ZHANG L. X.; ZHANG, C. L.; LI, W.J. Diversity of Crenarchaeota in terrestrial hot springs in Tengchong, China. **Extremophiles**. 2010.

STEIN, L. Genome annotation: From sequence to biology, **Nature Reviews Genetics** 2. 7:493-503, 2001.

STOUT, L.M.; BLAKE, R. E.; GREENWOOD, J. P.; MARTINI, A. M.; ROSE, E.C. Microbial diversity of boron-rich volcanic hot springs of St. Lucia, Lesser Antilles. **FEMS Microbiol Ecol.** 2009.

STRAATHOF, A. J.; PANKE, S.; SCHMID, A. The production of fine chemicals by biotransformations. **Current opinion in biotechnology**, 13(6), 548-556, 2002.

SCHWEDE, T. Protein modeling: what happened to the “protein structure gap”? *Structure*, 21(9), 1531-1540. 2013.

TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular Biology and Evolution**, 28, 2731-2739. 2011 <http://dx.doi.org/10.1093/molbev/msr121>

TEELING, H.; GLÖCKNER, F. O. Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective. **Briefings in bioinformatics**, 13(6), 728-742, 2012.

TANG, W.M.; TO, K.Y. Four X-chromosomal STRs and their allele frequencies in a Chinese population. **Forensic Sci Int.** 162(1-3):64-5, 2006.

TEIXEIRA, L. C.; PEIXOTO, R. S.; CURY, J. C.; SUL, W. J.; PELLIZARI, V. H.; TIEDJE, J.; ROSADO, A. S. Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. **ISME J**, 2010.

TORSVIK, V., ØVREÅS, L. Microbial diversity and function in soil: from genes to ecosystems. **Current opinion in microbiology**, 5(3), 240-245, 2002.

VICK, T. J.; DODSWORTH, J. A.; COSTA, K. C.; SHOCK, E. L.; HEDLUND, B. P. Microbiology and geochemistry of Little Hot Creek, a hot spring environment in the Long Valley Caldera. **Geobiology** 8: 140-154, 2010.

WANG, Y.; JIN, S.; WANG, M.; ZHU, L.; ZHANG, X. Isolation and Characterization of a Conserved Domain in the Eremophyte H⁺-PPiase Family. **PloS one**, 8(7), e70099, 2013.

WHITMAN, W. B.; COLEMAN, D. C.; WIEBE, W. J. Prokaryotes: The unseen majority. **Proceedings of National Academy of Sciences of United States of America**, v. 95, p. 6578-6583, 1998.

WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic acids research**, 35(suppl 2), W407-W410, 2007.

YOUNG, T. W.; KUHN, N. J.; WADESON, A.; WARD, S.; BURGESS, D.; COOKE, G. D. *Bacillus subtilis* ORF yybQ encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase? **Microbiology** 144:2563–2571, 1998.

ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **BMC bioinformatics** 5(4): 725-738, doi:10.1186/1471-2105-9-40, 2008.

ZYRYANOV, A. B.; VENER, A. V.; SALMINEN, A.; GOLDMAN, A.; LAHTI, R.; BAYKOV, A. A. Rates of elementary catalytic steps for different metal forms of the family II pyrophosphatase from *Streptococcus gordonii*. **Biochemistry** 43:1065-1074, 2004.

8- APÊNDICE

Tabela A1 - Resultados obtidos por homologia utilizando Blastx, PFame ProDom

ORFs	Função /N° de acesso	E- value	Ident. %	Superfamilies	ProDom
1	Hypoxanthine Phosphoribosyltransferase (WP_007573236)	1e-78	67	PRTases _ Type I	Glycosyltransferase Phosphoribosyltransferase (PD404971)
2	tRNA(Ile)-Lysidine Synthetase (YP_003393473)	2e-34	46	AANH_Like	ATP-Binding Nucleotide-Binding tRNA (PD003493)
3	Hypothetical Protein (WP_019639547)	1e-53	46.	Dehydrogenase, PQQ -Dependent	Putative Kinase Uncharacterized Lyase (PD886267)
4	N-Acetyltransferase GCN5 (YP_004922604)	5e-26	37	NAT_SF	Transferase (PD598458)
5	Purine-Nucleoside Phosphorylase (YP_643389)	1e-88	59	PNP_UDP_1	AMP Hydrolase (PD590616)
6	Lipid Kinase (WP_007784249)	3e-60	43	DAGK_cat	Kinase(PD604008)
7	Factor Regulator FecR (WP_007573104)	1e-74	50	SIR2	Deacetylase SIR2 (PD280204)
8	Thioesterase Superfamily Protein (YP_00370504)	1e-20	43	Hot_Dog	Thioesterase (PDB553B6)
9	Vancomycin Resistance Protein (YP_005330201)	1e-46	36	VanW	Uncharacterized (PDA4H3X1)
10	Enoyl-CoA Hydratase (WP_0174349640)	1e-47	53	Crotonase _ like	Hydratase/Isomerase (PDB1K0N7)
11	Short-Chain Dehydrogenase (WP_017434965)	5e-109	75	NADB_ Rossmann	Putative Dehydrogenase (PDA5N6N6)
12	Acyl-CoA Dehydrogenase (WP_00757311)	1e-83	65	Acyl-CoA	Acyl-CoA Oxidoreductase Domain (PDC2C2H4)
13	Acyl-CoA Dehydrogenase (WP_00757311)	5e-14	59	Acyl-CoA_dh_N	Acyl-CoA Domain (PD000558)
14	Acyl-CoA Dehydrogenase (YP_003393461)	1e-07	62	Acyl-CoA_dh_N	Unknown

Tabela 1: suplementar continuaç

ORFs	Função /N° de acesso	E- value	Ident. %	Superfamilies	ProDom
15	Membrane Protein (WP_018959726)	6e-19	43	Lysyl-tRNA Synthetase	Putative Uncharacterized Synthetase (PD014386)
16	NADH-Quinone Oxidoreductase Subunit B (WP_020472635)	1e-68	81	Oxidored_q6	Oxidoreductase (PD020213)
17	NADH- Ubiquinone/Plastoquinone Oxidoreductase chain 3 (YP_003369289)	2e-56	70	Oxidored_q4	Oxidoreductase (PDC48838)
18	N-Acetyltransferase (YP_001434269)	1e-43	51	NAT_SF	Transferase (PD485972)
19	Short-Chain Dehydrogenase (WP_020605594)	3e-125	67	NADB_Rossmann	Oxidoreductase Dehydrogenase/Reductase (PD332894)
20	Inorganic Diphosphatase (YP_004041784)	0.0	70	H ⁽⁺⁾ Translocating Inorganic Pyrophosphatase TEX_N	Hydrolase (PD852318)
21	RNA binding (YP_003368608)	0.0	76		Binding Domain S1 (PD491134)47
22	Hypothetical Protein (WP_002645780)	8e-176	57	DUF1501	Uncharacterized (PDA1F622)
23	Hypothetical Protein (YP_003369169)	0.0	80	DUF1501	Uncharacterized (PDA1F622)
24	RecQ FamilyATP-Dependent DNA helicase (YP_003371490)	0.0	62	DEXDc	Helicase Hydrolase ATP-Binding (PDA0Z8A2)

Tabela A 1: Sequências de PPase integral de membrana usadas para análise fenética.

Nome das Espécies	N° de acesso		
<i>naerofustis_stercorihominis</i>	ZP_02861667	<i>Polaribacter_irgensii</i>	ZP_01118223.1
<i>Anaerostipes_caccae</i>	ZP_02417489	<i>Pyrobaculum_aerophilum</i>	NP_559532.1
<i>Arabidopsis_thaliana</i>	AAA32754	<i>Rhodospirillum_rubrum</i>	AAC38615.2
<i>Arabidopsis_thaliana_(AVP2)</i>	AAC83018.1	<i>Rhodothermus_marinus</i>	ZP_04423801.1
<i>Bacteroides_pectinophilus</i>	ZP_0346300.1	<i>Robiginitalea_biformata</i>	YP_003196855
<i>Bdellovibrio_bacteriovorus</i>	NP_968591.1	<i>Salinibacter_ruber</i>	YP_444948.1
<i>Blautia_hydrogenotrophica</i>	ZP_03783250.1	<i>Streptomyces_coelicolor</i>	NP_627745.1
<i>Bradyrhizobium_diazoeficiens</i>	NP_771666.1	<i>Streckbrandtia_nassauensis</i>	ZP_04482946.1
<i>Caldvirga_maquilingensis</i>	YP_001540035		
<i>Caldvirga_maquilingensi</i>	YP_001540771		
	YP_001957756		
<i>Candidatus_Amoebophilus_asia</i>			
<i>ticus</i>			
<i>Capnocytophaga_sputigena</i>	ZP_03392450.1		
<i>Carboxydotherrnus_hydrogenof</i>	YP_359158.1		
<i>rmans</i>			
<i>Caulobacter_crescentus</i>	NP_420176.1		
<i>Cellulomonas_flavigena</i>	ZP_04367522.1		
<i>Chlamydomonas_reinhardtii</i>	CAC44451.1		
<i>Chlorobium_limicola</i>	YP_001943220		
<i>Chloroflexus_aurantiacus</i>	YP_001634923		
<i>Chthoniobacter_flavus</i>	ZP_03127672.1		
<i>Clostridium_botulinum_E3</i>	YP_001921624		
<i>Clostridium_tetani</i>	NP_781083.1		
<i>Clostridium_tetani</i>	NP_781083.1		
<i>Clostridium_thermocellum</i>	YP_001037489		
<i>Coprothermobacter_proteolyticu</i>	YP_002246899		
<i>s</i>			
<i>Cyanidioschyzon_merolae</i>	BAE19660.1		
<i>Dehalococcoides_ethenogenes</i>	YP_181498.1		
<i>Dehalococcoides_ethenogenes</i>	YP_181516.1		
<i>Denitrovibrio_acetiphilus</i>	ZP_03906694.1		
<i>Desulfobacterium_hafniense</i>	YP_521069.1		
<i>Desulfococcus_oleovorans</i>	YP_001530700		
<i>Desulfuromonas_acetoxidans</i>	ZP_01313190.1		
<i>Dethiobacter_alkaliphilus</i>	ZP_03729322.1		
<i>Dictyoglomus_thermophilum</i>	YP_002251142		
<i>Diplosphaera_colitermitum</i>	ZP_03727379		
<i>Dorea_longicatena</i>	ZP_01995168		
<i>Fervidobacterium_nodosum</i>	YP_001410846		
<i>Flavobacterium_johnsoniae</i>	YP_001193830		
<i>Gallionella_ferruginea</i>	ZP_04831980.1		
<i>Haliangium_ochraceum</i>	ZP_03880259.1		
<i>Halothermothrix_oreni</i>	YP_002509320		
<i>Heliobacterium_modesticaldum</i>	YP_001679270		
<i>Hyphomonas_neptunium</i>	YP_759877.1		
<i>Jonquetella_anthropi</i>	ZP_05861386.1		
<i>Kosmotoga_olearia</i>	YP_002939864		
<i>Leptospira_biflexa</i>	YP_001840784		
<i>Leptospira_interrogans</i>	NP_711652.2		
<i>Mesorhizobium_lotii</i>	NP_108517.1		
<i>Methanosarcina_mazei</i>	NP_632724.1		
<i>Methanospirillum_hungatei</i>	YP_503835.1		
<i>Methylacidiphilum_infernorum</i>	YP_001940861		
<i>Methylophilales_bacterium</i>	ZP_01551779.1		
<i>Moorella_thermoacetica</i>	YP_430205.1		
<i>Moorella_thermoacetica</i>	YP_429292.1		
<i>Nocardioideis_sp</i>	YP_921602.1		
<i>Oxalobacter_formigenes</i>	ZP_04577892.1		
<i>Parvimonas_micra</i>	ZP_02094589.1		
<i>Pelotomaculum_thermopropioni</i>	YP_001213266		
<i>cum</i>			
<i>Perkinsus_marinus</i>	EER08598.1		
<i>Plasmodium_falciparum</i>	AAG21366.1		
<i>Plasmodium_falciparum</i>	AAD17215.1		
<i>Plesiocystis_pacifica</i>	ZP_01910339.1		
<i>Polaribacter_irgensii</i>	ZP_01118223.1		
<i>Pyrobaculum_aerophilum</i>	NP_559532.1		
<i>Rhodospirillum_rubrum</i>	AAC38615.2		
<i>Rhodothermus_marinus</i>	ZP_04423801.1		
<i>Robiginitalea_biformata</i>	YP_003196855		
<i>Salinibacter_ruber</i>	YP_444948.1		
<i>Plasmodium_falciparum</i>	AAG21366.1		
<i>Plasmodium_falciparum</i>	AAD17215.1		
<i>Plesiocystis_pacifica</i>	ZP_01910339.1		