

**UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**ANÁLISE *in silico* DE UM NOVO “CLUSTER” DE PKS II
METAGENÔMICO**

Elisângela Soares Gomes

Bióloga

JABOTICABAL – SÃO PAULO – BRASIL

Fevereiro de 2011

**UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**ANÁLISE *in silico* DE UM NOVO “CLUSTER” DE PKS II
METAGENÔMICO**

Elisângela Soares Gomes

Orientadora: Profa. Dra. Eliana Gertrudes de Macedo Lemos

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

JABOTICABAL – SÃO PAULO – BRASIL

Fevereiro de 2011

G633a Gomes, Elisângela Soares
Análise *in silico* de um novo “cluster” de PKS II Metagenômico /
Elisângela Soares Gomes. – – Jaboticabal, 2011
xvi, 118 f. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2011
Orientadora: Eliana Gertrudes de Macedo Lemos
Banca examinadora: Gabriel Padilla Maldonado, Jesus Aparecido
Ferro
Bibliografia

1. PKS II. 2. Metagenoma. 3. Bioinformática estrutural I. Título. II.
Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 575.111

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 UM NOVO "CLUSTER" DE PKS II METAGENÔMICO

AUTORA: ELISÂNGELA SOARES GOMES

ORIENTADORA: Profa. Dra. ELIANA GERTRUDES DE MACEDO LEMOS

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

Profa. Dra. ELIANA GERTRUDES DE MACEDO LEMOS

Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. JESUS APARECIDO FERRO

Departamento de Tecnologia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. GABRIEL PADILLA MALDONADO

Departamento de Microbiologia / Instituto de Ciências Biomédicas - USP / São Paulo/SP

Data da realização: 23 de fevereiro de 2011.

DADOS CURRICULARES DO AUTOR

ELISÂNGELA SOARES GOMES – Nasceu em Londrina (PR, Brasil) no dia 05 de janeiro de 1985. Ingressou na “Faculdade de Ciências Agrárias e Veterinárias” – FCAV-UNESP/Jaboticabal em março de 2005, no curso de Ciências Biológicas (modalidade Bacharel), o qual foi concluindo em dezembro de 2008. No ano de 2009 ingressou na modalidade de Licenciatura do curso de Ciências Biológicas, com término em dezembro de 2009. Durante os cinco anos de graduação realizou atividades de estágio no Laboratório de Bioquímica de Micro-organismos e Plantas (LBMP), sob orientação da Profa. Dra. Eliana Gertrudes de Macedo Lemos. No período de 2005-2006 realizou estágio voltado para o cultivo *in vitro* e conservação de *Orchidaceae*, acompanhando o trabalho do então graduando Renato Fernandes Galdiano Júnior, participando de trabalhos envolvendo o cultivo *in vitro* de sementes de *Cattleya walkeriana*, cultivo por gêmula floral de *Phalaenopsis*, cultivo por calos de raízes de *Cattleya walkeriana* e documentação por Microscopia Eletrônica de Varredura de estágios iniciais de germinação de *Cattleya walkeriana*. No período de julho de 2006 à julho de 2007 acompanhou a então doutoranda Viviane Schuch em seu trabalho de prospecção de genes envolvidos na síntese de antibióticos do sistema PKS I (Modular), envolvendo a construção de bibliotecas metagenômicas e busca por sequências conservadas das mega-síntases, tendo oportunidade de aprendizagem de técnicas de biologia molecular e uso de ferramentas de bioinformática. No período de agosto de 2007 à julho de 2009 desenvolveu uma atividade de prospecção de clones metagenômicos abrigando agrupamentos gênicos relacionados a síntese de antibióticos policetílicos aromáticos (PKSII) com Bolsa de auxílio FAPESP (processo 07/53708-0). Em março de 2010 ingressou no curso de Mestrado em Microbiologia Agropecuária (FCAV-UNESP) realizando, como atividade de pesquisa, a predição funcional *in silico* de genes envolvidos na síntese de antibióticos policetílicos aromáticos, conforme o descrito no presente trabalho, com Bolsa de Auxílio do Programa oferecida pelo CNPq (processo 130454/2010-9). Em comum acordo com a orientadora, a defesa do título de Mestre foi antecipada para fevereiro de 2011, e atualmente a aluna é ingressa no curso de Doutorado em Microbiologia Agropecuária (FCAV-UNESP).

EPÍGRAFE

*“Jamais considere seus estudos como uma obrigação,
mas como uma oportunidade invejável para
aprender a conhecer a influência libertadora da beleza
do reino do espírito, para seu próprio
prazer pessoal e para proveito da comunidade
à qual seu futuro trabalho
pertencer”.*

Albert Einstein

DEDICATÓRIA:

Ao Prof. Dr. Charles Richard Hutchinson[†], pelo seu importante legado para a pesquisa em PKSs, resultado de toda uma vida de empenho e dedicação à pesquisa.

À sociedade, pela e para qual esse trabalho foi possível.

AGRADECIMENTOS:

- ★ À minha orientadora, Profa. Dra. **Eliana G. de Macedo Lemos**, pelo constante incentivo, pela sabedoria das palavras diante de cada impasse, pela paciência e pela confiança em mim depositada;
- ★ À minha família: **Luiza** (mãe), **José Cícero** (pai), **José Clemente** (pai de coração), **Cristiano** e **Jaqueline** (Irmãos), pela estrutura e estabilidade emocional.. já viram uma casa sem alicerce?
- ★ À doutoranda **Viviane Schuch**, pela direta contribuição pessoal, profissional e intelectual durante o percurso deste trabalho e minha jornada no LBMP, bem como pela amizade;
- ★ Aos demais pesquisadores que contribuíram para esse trabalho: Prof. Dr. **Gabriel Padilla Maldonado**, Prof. Dr. **Jesus Aparecido Ferro**, Profa. Dra. **Silvana Giulliaty**, Dr. **Luciano Takeshi Kishi**, Dra. **Lúcia M. Cararetto Alves**, Dr. **Maurício Egídio Cantão**;
- ★ Aos amigos com os quais eu tive a honra de trabalhar direta e indiretamente, representados nas pessoas dos doutores **João Carlos Campanharo**, **Jackson Marcondes de Souza**, **Simone C. Pick**, **Ester Wekert**, **Silvana Pompeia Do Val**, **Viviane Vieira**, **Renata Andrade**; **Tereza Cristina Castellane**; **Eliamar N. Pedrinho**; doutorandos: **Renato F. Galdiano Jr.**, **Gisele Regina Rodrigues**; **Wellington Moreira**, **Karla Cristina Stropa**; **Fernanda Paganelli**, **Mauricio R. Demitrova**, **Douglas Paixão**, **Micheli**, das mestrandas: **Mariana Rangel**, **Thais Maester**, **Erica Mendes Lopes**, **Andressa**, **Ludmila**, **Maira**, aos graduandos **Carlos**, **Wellington**, **Milena**, **Rosmeriana**, **Micheli Persona**, **Maria Luíza**, **Helena**, enfim: **à todos os meus amigos do LBMP**, pela ajuda de sempre, pelo ambiente amistoso, propiciando o crescimento mútuo pessoal, afetivo e profissional, e pela amizade de sempre, sem a qual não existe a satisfação pessoal;
- ★ Aos meus professores e professores doutores, bem representados (porém não restritos à esses) nas pessoas da **Eliana**, **Manuel Victor**, **Lúcia**, **Jesus**, **Maria Inês**, **Jackson**, **Wagner**, **Rosemary**, **Ferraudo**, **Danízio**, **Roberto**, **Monteiro**, **Durvalina**, **Célia**, **Barbanti**, **Silvana**, **Gilberto** e, porque não, **Maristela** (quinta série do ensino fundamental), por me ensinarem à aprender à “aprender”
- ★ Aos meus amigos de turma da graduação (**Bio05**), representados nas pessoas **Jaqueline**, **Gustavo**, **Andreza**, **Rodolfo**, **Fernanda**, **Jhoane**, **Aline**, **Ariadne**, **Analuê**, **Micaela**, **Luciana**, e **demais** que estiveram comigo desde os primeiros passos dessa caminhada;
- ★ À **Fapesp**, **CNPq**, e ao **Programa de Microbiologia Agropecuária**, pelo suporte à essa pesquisa.

★★

SUMÁRIO

	Página
SUMÁRIO	viii
ABREVIATURAS	xi
ÍNDICE DE FIGURASxiii
RESUMO	xv
SUMMARYxvi
CAPÍTULO 1 - Considerações Gerais1
1.1. INTRODUÇÃO1
1.2. OBJETIVOS DO TRABALHO4
1.3. REVISÃO DE LITERATURA5
1.3.1. Metagenômica5
1.3.2. Estudos em Metagenômica realizados no LBMP7
1.3.3. Bioinformática e Filogenômica9
1.3.4. Metabólitos secundários10
1.3.5. Actinomicetes: riqueza de metabólitos secundários11
1.3.6. O gênero <i>Streptomyces</i> e a produção de metabólitos secundários15
1.3.7. Análise do setor farmacológico e perfil das organizações envolvidas na descoberta de novas drogas19
1.3.8. Prospecção de antibióticos empregando Biologia Molecular24
1.3.9. Policetídeos27
1.3.10. Tipos de PKS e Enzimologia dos Grupos29
1.3.11. Origem das PKSs33
1.3.12. As PKS II38
1.3.13. Biossíntese combinatória40
1.3.14. Breve revisão de Estruturas de Proteínas43
1.3.15. Modelagem enzimática in silico de estruturas 3D e seu “Assessing”45
CAPÍTULO 2 – Análise funcional por homologia de um “cluster” de PKSII metagenômico48

	ix
2.1. Resumo	48
2.2. Introdução.	49
2.3. Material e Métodos	50
<i>Biblioteca</i>	<i>50</i>
<i>Prospecção de agrupamentos gênicos de PKSII</i>	<i>51</i>
<i>Subclonagem, sequenciamento do cosmídeo escolhido e análises de bioinformática</i>	<i>52</i>
<i>Análises filogenéticas</i>	<i>53</i>
<i>Comparação de “clusters”</i>	<i>54</i>
2.4. Resultados E Discussão	55
<i>Prospecção de genes para a PKS II</i>	<i>55</i>
Inferências de função por homologia.	56
ORFs para a <i>pks mínima</i>	56
<i>Genes de unidades enzimáticas de modelagem intermediária da cadeia</i>	<i>58</i>
<i>Genes relacionados a modificação pós-síntese da cadeia policetídica</i>	<i>59</i>
<i>Genes relacionados a regulação transcricional e transporte de substâncias</i>	<i>61</i>
<i>Genes possivelmente envolvidos com PKS II e degradação de compostos lignínico-celulósicos e outros genes.</i>	<i>62</i>
<i>Análises filogenéticas Comparação de “clusters”</i>	<i>66</i>
2.5. Conclusões.	70
CAPÍTULO 3 – Predição estrutural <i>in silico</i> por homologia de uma PKS Mínima	
Metagenômica	72
3.1. Resumo	72
3.2. Introdução.	73
3.3. Material e Métodos	74
<i>Análises de modelagem estrutural.</i>	<i>74</i>
3.4. Resultados e Discussão	74
<i>Modelagem 3D in silico e análises de “assessing”</i>	<i>74</i>
3.5. Conclusões	88

CAPÍTULO 4 – Análise in silico de O-metil-transferase de interesse biotecnológico <i>acessada por metagenômica</i>89
4.1. Resumo89
4.2. INTRODUÇÃO	90
4.3. MATERIAL E MÉTODOS.	93
<i>Biblioteca</i>	93
<i>Prospecção de genes de PKS</i>93
<i>Subclonagem, sequenciamento do cosmídeo escolhido e análises de bioinformática</i>94
<i>Análises de modelagem estrutural</i>	95
4.4. RESULTADOS E DISCUSSÃO96
<i>Identificação da sequência metagenômica por homologia</i>	96
<i>Modelagem 3D in silico e análises de “assessing”</i>	97
4.5. Conclusões.101
CAPÍTULO 5 – Implicações	101
6. REFERÊNCIAS.103

ABREVIATURAS

3-K-ACP-REDUC.....	3-Ketoacil ACP-Redutase;
3-K-ACP-SYN.....	3-Keto-acil ACP-Síntase.
ABC 2.....	celbiose ABC transporter permease protein/ "transport system inner membrane protein";
ABC 3.....	celbiose transportador permease;
ABC1.....	ABC transportador celbiose: proteína ligante de soluto;
ACD.....	Acil-CoA Deidrogenase;
ACIL-TRANS.....	Acil-transferase;
ACP.....	Acyl Carrier Protein;
CHITINASE.....	Hidrolase/quitinase;
CLF.....	<i>Chain lenght factor</i> ;
CLP.....	ATP-dependent Protease Homóloga;
CYC 2/ARO	Policetideo Ciclase I/Aromatase;
CYC 3	Policetideo Ciclase II;
CYC1.....	Policetideo Ciclase I;
D-HTH.....	HTH Domínio proteína/plu4195;
DH.....	Dehidratase;
FabH.....	" β -Ketoacyl Acyl Carrier Protein Synthase III"
GLI-HIDROL	Hidrolase Beta-Glucosidase;
GLI-TRANS.....	Glicosil-Transferase;
HIPOT	Proteína Hipotética/Proteína Secretada;
HYDR	Hydroxylase;
KR.....	Ketoreductase;
KSA	β -cetoacil sintase α ;
MONOX	Mono-oxigenase;
MT-TRANS.....	Metil-Transferase;
MULT-COPP	Multicopper oxidase;
O-METH	O-Methyl transferase;
OMT-M.....	O-Methyl transferase Metagenômica;
OXD.....	Oxidoreductase;
P-450	Citocromo P450;

PEP-TRANSP.....	Peptideo Transportador;
PEPTIDASE.....	Peptidase/Sortase;
REG.....	Regulador transcricional;
SAN-1.....	Acyl-CoA descarboxilase;
SAN-2.....	Acil-ACP Tioesterase;
SAN-3.....	Mono-oxigenase formadora de quinona;
TRANSP/REG.....	citação genérica para genes envolvidos em transporte/regulação;
UNKNOW.....	Proteínas Desconhecidas ou não citados na fonte de referência;

ÍNDICE DE FIGURAS

Fig.1 - Representação diversidade de <i>Actinomycetes</i> (GENILLOUD et al., 2010)	12
Fig.2 - Alguns novos compostos de origem da classe <i>Actinomycetes</i> descobertos em 2010 (GENILLOUD et al., 2010)	13
Fig.3 - Representação diversidade (16S rDNA) de <i>Actinomycetes</i> acessados por isolamento sistemático (GENILLOUD et al., 2010)	14
Fig.4. Representação gráfica espacial do códon de uso de organismos frequentemente estudados (GUSTAFSSON et al. 2004)	17
Fig.5 - Distribuição por família química e por estágio de pesquisas para novas drogas (HAMAD 2010)	20
Fig.6 - Médias obtidas para o tempo de retenção de novas moléculas de acordo com a área patológica, obtidos no período de 2003 à 2007 (KAITIN, 2010)	21
Fig.7 - Contribuições da iniciativa pública e privada para a descoberta de novos fármacos para o período entre 1998 e 2007 (KNELLER, 2010)	22
Fig.8 - Avaliação de novos medicamentos lançados no Brasil no período de 1998 à 2001 (WADT, 2003).	23
Fig.9 - Árvore filogenética construída de sequências de aminoácidos de <i>KSα</i> de <i>Actinomicetes</i> (METSÄ-KETELÄ et al., 1999)	26
Fig.10 - Estrutura de policetídeos representativos, incluindo bioativos proeminentes. (WEISSMAN, 2009)	28
Fig.11 - Esquema de reações e dinâmica catalítica das PKSs (WEISSMAN, 2010) . . .	31
Fig.12 - Esquema Genérico das reações para biossíntese tanto de ácido graxo quanto de policetídeos (WEISSMAN, 2009)	34
Fig.13 - Comparação filogenética entre FAS e PKSs (JENKE-KODAMA e DITTMANN, 2009)	36
Fig.14 - Esquema da estrutura primária da ribonuclease (SWISS MODEL)	44

Fig.15 - Esquema representativo de estruturas protéicas..	45
Fig.16 - Amplificação das amostras compostas dos clones metagenômicos.	55
Fig.17 - Mapa da distribuição de ORFs (1-22: Blastp) e Domínios conservados (A-U: PRODOM) obtidos para o “Cluster” metagenômico	56
Fig.18 - Comparação filogenética de genes para KSA de vários “clusters” de PKS disponíveis na literatura e a ORF Metagenômica.	68
Fig.19 - Análise comparativa entre o <i>contig</i> metagenômico B5-37 e diversos “cluster” para PKS II.	69
Fig.20 - Predição funcional para o modelo e a ORF15.	78
Fig. 21 - Comparação da estrutura 2D entre a ORF15 e o molde 1TQYA.	79
Fig.22 - Avaliação da qualidade da estrutura 3D por meio do Qmean.	80
Fig.23 - Modelo 1TQY na forma octomérica (dímeros CLF/ACP).	82
Fig.24 - Comparação entre o estrutura de um dos dímeros 1TQY e os resultados obtidos para o “docking” entre a subunidade metagenômicas KSA e CLF	83
Fig.25 - Comparação entre as estruturas da ACP 1NQ4A e o modelo preditivo para a ACP metagenômica	84
Fig.26 - Análise de compatibilidade por simulação de “docking” entre KSA Metagenômica contra a CLF Metagenômica, a CLF 1tqyH, e CLF 1j3nA.	85
Fig.27 - Predição estrutural para o molde e a OMT-M com análise de erros de estrutura pelo Qmean	99
Fig.28 - Predição estrutural para o molde e a OMT-M, destacando as cadeias enzimáticas do molde 2IP2, com recursos Wizard do MacPymol.	100

ANÁLISE *in silico* DE UM NOVO “CLUSTER” DE PKS II METAGENÔMICO

RESUMO – As PKSs tipo II são complexos enzimáticos envolvidos na produção de policetídeos. Estes são metabólitos secundários que possuem papel-chave no desencadeamento das respostas bioquímicas que levam à diversificação fisiológica frente a fatores adversos, fazem parte da composição de pigmentos de esporos microbianos, além de comporem a estrutura química de importantes antibióticos explorados pela indústria farmacêutica. Foi feita a prospecção de novos “clusters” gênicos para PKSs tipo II em uma biblioteca metagenômica de 9.320 clones, através de amplificação por PCR com “primers” degenerados para uma região gênica conservada. Dentre três clones encontrados, um foi sequenciado pela estratégia de *shotgun*, resultando em um *consensus* de 22.988 pares de bases. Foi feita a predição funcional de ORFs por homologia obtida com a ferramenta Blast (NCBI) e também a identificação de domínios enzimáticos pelo PRODOM e PFAM. Dentre os resultados da predição por homologia de domínios enzimáticos foi possível identificar ORFs homologas à: reguladores gênicos (LacI e LuxR); carboidrases (Lacases, β -glucosidases, Quitinases/Celulases); transferases (O-metil-transferases, Acil-transferases); pks mínima; ciclases, aromatases, hidroxilases, mono-oxigenases e transportadores ABC. Alguns “clusters” enzimáticos podem ainda atuar em degradação de celulose e lignina. As enzimas candidatas ao *core* biossintético do “cluster” (pks mínima) foram submetidas à modelagem de estrutura tridimensional *in silico*. Os modelos 3D foram avaliados quanto a acurácia, utilizando as ferramentas: Verify3D, Qmean, Procheck, Rampage e Swiss-Model. Uma vez validados, as KSA e CLF foram submetidas a simulações de “docking” enzimático.

Palavras-chave: PKSII, Metagenoma, Bioinformática estrutural, antibióticos. Degradação de lignina/celulose

***In silico* ANALYSIS OF PKS II A NEW "CLUSTER" OF PKS II metagenomic**

SUMMARY - The type II PKSs are enzyme complexes involved in the production of polyketides. These are secondary metabolites that have key role in triggering the biochemical responses that lead to physiological diversification against adverse factors are part of the pigment composition of microbial spores, and compose the chemical structure of most antibiotics exploited by the pharmaceutical industry. Some clusters could act in enzymatic degradation of cellulose and lignin. It was made a prospection for new clusters gene for type II PKSs in a metagenomic library of 9320 clones by PCR amplification with degenerate primers for a conserved gene region. Among three clones found, one was sequenced by subcloning and shotgun strategy, resulting in a consensus of 22,988 base pairs. It was made the prediction of functional homology ORFs by Blast (NCBI) and also identifying areas for enzymatic Prodom and PFAM. Among them were found homologous to: ABC transporters; minimal PKS, cyclasea; Aromatasea; Hydroxylases; Mono-oxygenases; O-methyl-transferases; Multicopper oxidase/laccase, β -glucosidase, chitinase/ Cellulases; Peptidases; Acyl-transferases, in addition of gene regulators similar to LacI and LuxR. The biosynthetic enzymes candidates for the *core* of the cluster (minimal pks) underwent three-dimensional structure modeling in silico. The 3D models were evaluated for accurate, using the tools of 'Assessing "Verify3D, Qmean, Procheck, Rampage and Swiss-Model. Once validated, the KSA and CLF were subjected to simulated docking enzyme between them.

Key-Words: PKSII, Metagenomics, , Structural Bioinformatics, Antibiotics, Degradation of lignin / cellulose.

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1.1. INTRODUÇÃO

Os policetídeos são originalmente produtos do metabolismo secundário de algumas espécies de bactérias, fungos, plantas, e mesmo microalgas como dinoflagelados, e compõem a estrutura química de muitos fármacos: antibacterianos, antifúngicos, antineoplásicos, imunossuppressores, agentes redutores de colesterol, analgésicos, neurotoxinas, antiparasitas e determinantes de virulência (RIX et al., 2002). Tal versatilidade é refletida em um faturamento anual superior a U\$ 20 bilhões, chegando à mais de um terço do total das novas drogas implementadas no período de 2005 à 2007 (WEISSMAN, 2009). Embora os números sejam significativos, eles não refletem uma abundância sobressalente na disponibilidade de novas moléculas, refreada pelo processo lento de obtenção de formulações farmacológicas finais, contrastando com plasticidade de adaptação dos micro-organismos patogênicos (HAMAD, 2010).

Impulsionada pelo progressivo declínio da recuperação de novos compostos naturais decorrente de prospecção tradicional, intensa, desde a descoberta da estreptomicina em 1950, uma alternativa surgiu com a Biossíntese Combinatória de Policetídeos em 1985, nos primeiros passos do chamado “desenho racional” de novos fármacos (HOPWOOD et al. 1985; RIX et al., 2002). Tal abordagem persiste, e com o advento da Metagenômica, tornou-se ainda mais poderosa, diante do acesso à microbiota de forma não restrita à escassa parcela de 1% cultivável em laboratório (BALTZ, 2006; COMMITTEE ON METAGENOMICS, 2007).

A despeito da prospecção por novos “clusters” biossintéticos de compostos policetídeos ter em vista, em sua maioria, a busca por metabólitos com atividade farmacológica, este grupo não se restringe a estes. É conhecido que além de atividade antimicrobiana, os compostos policetídeos podem estar relacionados à constituição de pigmentos de esporo. Os mecanismos envolvidos em sua síntese também desencadeiam sinais bioquímicos que induzem a diferenciação fisiológica adaptativa

aos processos de resistência às condições ambientais adversas (BODE & MULLER, 2005; DAVIES, 2006; CHALLIS, 2008; JENKE-KODAMA & DITTMANN, 2009). Além disso, ainda há muito o que descobrir sobre estas vias, como ilustrado pela recente indicação de um desvio funcional de uma PKS II, originalmente envolvida na produção de melanina, para uma via de degradação de compostos lignínico-celulósicos (HSIAO & KIRBY, 2008). Sob este aspecto, o cenário atual de sequenciamento massivo de genomas e o acesso ao material genético total de microbiotas (Metagenoma) abre um horizonte eclético para a descoberta de novas vias de policetídeos, desde que bem respaldados por estratégias mais efetivas de prospecção e caracterização de novas rotas biossintéticas, chegando a pistas para sua função ecológica, ao mesmo tempo em que permitam diagnosticar eventuais aplicações para a Engenharia Molecular.

Neste trabalho, empregou-se oligonucleotídeos degenerados para uma região gênica conservada de PKS II para prospectar “clusters” em uma biblioteca metagenômica, seguido de sequenciamento do inserto, visando encontrar novos genes potencialmente úteis para a biotecnologia. Para o estudo de uma sequência encontrada foram identificadas no inserto as ORFs que apresentaram melhor resultado de homologia (ORFfinder e BlastP), e o “cluster” gênico resultante foi submetido a duas abordagens distintas de caracterização, sendo a primeira com base em homologia, respaldada por parentesco filogenético e arquitetura de “cluster”; e a segunda fundamentada no estudo de estrutura tridimensional protéica.

A primeira abordagem (Capítulo 2) consistiu de uma confirmação de homologia e correção de anotação de códons de iniciação com base em ortólogos conhecidos (BlastX, BlastP, BlastN e InterPro), seguido por análise de domínios enzimáticos (Prodom, PFam) contrastando os dados obtidos com a literatura pertinente ao estudo de enzimas de PKSII, obtendo-se hipóteses de homologia mais plausíveis e detalhadas para as ORFs. Em seguida, explorou-se o emprego de análises filogenéticas de uma subunidade enzimática chave no processo de escolha de substratos (KSA) das vias de PKSII, a qual permite, frequentemente, classificar funcionalmente vias de policetídeos em rotas de pigmento ou antibióticos (METSÄ-KETELÄ et al., 1999). Complementando o estudo, foi feita a análise da arquitetura do “cluster” metagenômico, comparando-o a

agrupamentos envolvidos com síntese de pigmentos e aos de síntese de antibióticos, enfatizando a composição e a disposição física dos genes.

A segunda abordagem, por sua vez, envolveu a síntese por homologia e validação de modelos tridimensionais para enzimas do “cluster” metagenômico. Como objetos de estudo, escolheu-se trabalhar o *core* biossintético *pks mínima* e a ORF22, relacionada à enzima **O-metil-transferase** de modificação da cadeia policetílica após a sua formação.

Para o estudo da *pks mínima* do “cluster” metagenômico (KSA, CLF e ACP) foram feitas análises envolvendo simulação de interação por superfície de contato entre subunidades protéicas (“Docking”), e identificação de diferenças nas suas estrutura 2D e 3D por comparação visual, buscando diagnosticar a possibilidade de emprego da mesma em uma estratégia de biossíntese combinatória.

Já para o estudo da O-metil-transferase, optou-se por realizar a comparação entre o modelo gerado e algumas estruturas disponíveis no PDB, elucidadas por métodos de modelagem clássica. O objetivo deste trabalho foi verificar o potencial desta enzima para a biossíntese combinatória para a geração de novos policetídeos.

Os resultados destas abordagens foram estruturados na forma de artigos científicos, obedecendo-se a organização textual proposta pelo Programa de Microbiologia Agropecuária (UNESP-Jaboticabal), para o modelo de Dissertação em capítulos.

1.2. OBJETIVOS DO TRABALHO

OBJETIVO GERAL:

- Prospecção de novos genes relacionados à família de PKS II, empregando uma Biblioteca Metagenômica;

OBJETIVOS ESPECÍFICOS:

- Realizar a predição funcional de ORFs por filogenômica, visando evidenciar genes de importância biotecnológica;
- Fazer a modelagem estrutural enzimática *in silico* por homologia de alguns dos prováveis produtos de expressão das ORFs, buscando relacioná-las à enzimas previamente caracterizadas experimentalmente;
- Analisar a compatibilidade de modelos de estruturas 3D e modelos obtidos por abordagens experimentais, como estratégia de confirmação da homologia inferida sobre as sequências nucleotídicas e de aminoácidos;
- Simular a recombinação entre ORFs metagenômicas e entre estas e modelos obtidos na literatura através de análise de interação proteína-proteína (“Docking”), buscando simular o uso das mesmas em biossíntese combinatória.

1.3. REVISÃO DE LITERATURA

1.3.1. Metagenômica:

Progressos na ecologia molecular microbiana demonstram que a diversidade genética de micro-organismos na natureza é muito maior que a refletida nas coleções e bancos de linhagens, estimando que apenas uma pequena parcela destes, em torno de 1%, é cultivável pelas técnicas padrões em meios de cultura (AMANN, LUDWING e SCHLEIFER , 1995; COMMITTEE ON METAGENOMICS, 2007).

A grande maioria dos micro-organismos possuem necessidades complexas demais para serem reproduzidas em ambientes artificiais, como os casos de indivíduos que apresentam sinergismos/sinetrofismos com outros da comunidade, carecendo de suplementação nutricional de aminoácidos, carboidratos não usais, e outros fatores de crescimento, e toda uma conjectura de fatores físico-químicos produzidos pelo metabolismo de outros micro-organismos, como pH e umidade ideais, redução de resíduos metabólicos nocivos, criação de ambiente anaero- ou aeróbicos, dentre muitas outras interações.

Metagenômica é uma metodologia que se baseia na análise genética e funcional da microbiota de forma independente de cultivo, interagindo diretamente sobre o “pool” de moléculas de ácidos nucléicos do ambiente. Em uma de suas aplicações, a metodologia faz uso da amplificação por PCR de regiões gênicas conservadas, seguido da clonagem e sequenciamento dos “amplicons”, criando as chamadas bibliotecas de diversidade gênica (COMMITTEE ON METAGENOMICS, 2007; GOMES et al., 2010).

Algumas regiões altamente conservadas ao longo de todos os genomas, como as regiões do DNA correspondentes a síntese do RNA ribossomal (rDNA e rRNA, respectivamente), sobretudo as regiões rDNA 16S (procariotos) e rDNA 18S (eucariotos) permitem classificar os organismos. Em virtude da extrema importância dessas moléculas para a manutenção funcional das células, suas sequências são altamente conservadas, sendo possível inferir o grau de parentesco de dois ou mais indivíduos através da comparação das diferenças acumuladas, lentamente, durante a

evolução das espécies, uma vez que mudanças muito bruscas e extensas, frequentemente, levam a deleção do organismo, enquanto que as adquiridas são dificilmente perdidas e serão compartilhadas por todos os seus descendentes mais próximos, permitindo rastrear a evolução.

Estudos empregando a amplificação de 16S rDNA e 18S rDNA de amostras de DNA extraídas diretamente do ambiente estão alterando dramaticamente as definições filogenéticas tradicionalmente aceitas dos principais grupos microbianos (BULL, WARD, & GOODFELLOW, 2000).

Outras regiões gênicas permitem também rastrear essa diversidade, sendo muitas vezes empregadas para verificar a riqueza para um grupo específico de enzimas ou reguladores, dentre outras proteínas presentes em uma dada comunidade, sendo importantes para se inferir o papel ecológico exercido pelas mesmas na manutenção do ecossistema estudado, bem como identificar ambientes ricos em enzimas de interesse biotecnológico.

Uma outra abordagem metodológica em metagenômica é a extração de DNA íntegro e destituído de impurezas, e clonagem deste material em vetores que comportem grandes insertos, tais como BACs e YACs (Cromossomos Artificiais Bacterianos e de Levedura, respectivamente), além de cosmídeos e fosmídeos, criando bibliotecas metagenômicas para a prospecção de genes de interesse ecológico e/ou biotecnológico, possibilitando a recuperação de vias biossintéticas totalmente desconhecidas. Um dos grandes entraves para a aplicação da técnica é a obtenção deste DNA de alta qualidade, devido a contaminação do mesmo com materiais ambientais coextraídos, como os ácidos fúlvicos e húmicos.

Atualmente existem alguns kits (ex.: *kit Ultra Clean Mega Soil DNA* da **Mo Bio**) e mesmo diversificados métodos de extração não comerciais (ex.: LILES et al., 2008) que permitem burlar em parte essas dificuldades, contudo a natureza físico-química do ambiente influi diretamente para o desempenho dos mesmos, carecendo de adaptações a cada situação estudada (MASSINI, 2009; SCHUCH, 2007).

A utilização da metagenômica como abordagem de busca por genes de interesse amplia em grande escala a perspectiva de sucesso destas estratégias, uma vez que

abrange uma diversidade não cerceada pela possibilidade ou não de cultivo laboratorial. Visto que a ampla diversidade microbiana do solo constitui-se de um enorme “pool” genético e biológico que pode ser explorado para a descoberta de novos genes, vias metabólicas inteiras e seus produtos (COWAN, 2005; BALTZ, 2006), a metagenômica oferece um caminho para acessar a diversidade microbiana de um ambiente em sua totalidade.

1.3.2. Estudos em Metagenômica realizados no LBMP

Em um dos primeiros trabalhos em Metagenômica realizados no Laboratório de Bioquímica de Micro-organismos e Plantas (LBMP), Departamento de Tecnologia, UNESP Campus de Jaboticabal, SILVEIRA e colaboradores (2006) compararam amostras ambientais de DNA obtidas de duas áreas: sob Arboreto de Eucaliptos e sob Floresta Nativa, com objetivo de verificar a biodiversidade das comunidades microbianas, através de estudos de DNA metagenômico de cada amostra, por Biblioteca de sequências de amplificação por PCR do 16S rDNA. Os resultados demonstraram que 72% das sequências analisadas eram novas, ou seja, vinham de bactérias que não haviam sido sequenciadas ainda, indicando uma importante fonte de sequências gênicas potencialmente úteis como ferramentas biotecnológicas. A composição da microbiota diferiu nos dois ambientes, sendo que na área sob Floresta Nativa, os micro-organismos pertencentes aos filos *Bacteroidetes* e *Planctomycetes*, apareceram em maior quantidade, enquanto na área sob Arboreto de Eucaliptos, predominaram as sequências classificadas como *Actinobacteria*, *Firmicutes* e *Verrucomicrobia*.

Um dado interessante foi a observação de que a população de actinomicetos foi maior em área sob Arboreto de Eucaliptos. Esses micro-organismos são de grande importância para a biotecnologia, como produtores de moléculas bioativas, e indicam que este solo pode ser uma fonte promissora para a prospecção de novos genes relacionados à produção de antibióticos (SILVEIRA *et al.*, 2006).

A biblioteca metagenômica utilizada nesse trabalho foi construída previamente (SCHUCH, 2007) com DNA metagenômico de elevado tamanho molecular isolado de solo coletado sob arboreto de eucaliptos, motivada pelos resultados de diversidade obtidos anteriormente. O solo foi coletado em um bosque plantado em fevereiro de 1969, em uma área pertencente ao Campus da Faculdade de Ciências Agrárias e Veterinárias, UNESP – Jaboticabal, estado de São Paulo, Brasil. A biblioteca possui 9.320 clones e foi construída em vetor cosmidial, sendo que o tamanho dos insertos clonados possuem em torno de 20 à 45 kb. Nessa biblioteca foram feitas a avaliação da presença de genes envolvidos na biossíntese de antibióticos do sistema PKS I (SCHUCH, 2007) e PKS II (GOMES, 2008); Lipases (PEREIRA, 2010), dentre muitos outros ainda não concluídos.

Além da biblioteca de solo sob serrapilheira de eucaliptos, foram construídas pela equipe do LBMP bibliotecas de solo de mangue, consórcio degradador de petróleo e solos sob palhada de cana-de-açúcar, sendo que outras estão em fase de construção.

Esta coleção tem sido explorada na busca por diversos genes de interesse industrial, tais como catalases, amilases, lipases, peptidases, celulases, vias de antibióticos, xilose-isomerases, fosfatases, dentre outras, empregando diferentes abordagens de prospecção, como o uso de sondas de hibridização, oligonucleotídeos iniciadores degenerados e ensaios bioquímicos.

Da mesma forma foram construídas outras bibliotecas de diversidade (16S rDNA, ITS), como as obtida para Lodo Ativado, Solo de Floresta Nativa, Monoculturas Comerciais e outras. A diversidade acessada será empregada para a construção de um “ecochip”, que poderá futuramente ser utilizado para identificar rapidamente organismos presentes em diferentes habitats, possibilitando estudos comparativos entre comunidades sob as variáveis de tempo e espaço.

1.3.3. Bioinformática e Filogenômica

A partir da segunda metade da década de 1990, com o aparecimento dos sequenciadores automáticos de DNA, ocorreu uma explosão na quantidade de sequências nucleotídicas a serem armazenadas e analisadas, demandando um aperfeiçoamento constante de recursos computacionais. Essa demanda criou uma interface entre as Ciências da Computação e a Biologia Molecular, bem como as áreas da Estatística, Matemática e Engenharia de Softwares, que fez emergir um novo campo da ciência: a Bioinformática (WWW.PROSDOCIMI, 2009).

Atualmente a bioinformática é uma ferramenta indispensável para processar o imenso volume de dados gerados por cada vez mais sofisticadas e elegantes abordagens metodológicas para acesso à diversidade, transformando-os em informações compreensíveis para a comunidade científica.

Dentre as ferramentas de bioinformática mais difundidas para análise de sequências, tem-se o Blast: “Basic Local Alignment Search Tool”, uma ferramenta de busca de similaridade entre sequências biológicas, tanto de DNA quanto de aminoácidos. O programa Blast é uma variação do algoritmo SMITH e WATERMAN (1981). O algoritmo Blast permite a comparação entre sequências de cinco formas diferentes: Blastp, Blastn, Blastx, tBlastn e tBlastx; conforme a necessidade do usuário e a natureza do banco de dados a ser consultado (EMBRAPA, 2007).

O programa Blastp compara sequências de aminoácidos de entrada (“query”) contra um banco de sequências de proteínas (“subject”). O programa Blastn, por sua vez, compara sequências de nucleotídeos de entrada contra um banco de entrada de sequências de nucleotídeos. Já o Blastx, equipara a sequências de nucleotídeos de entrada, traduzida para todas as fases de leitura possíveis, contra um banco de dados de sequências de proteínas. Por conseguinte, tem-se o programa tBlastn, que realiza a comparação de uma sequência de aminoácidos de entrada contra um banco de dados de sequências de nucleotídeos, traduzidas para todas as fases de leitura possíveis. O quinto programa do algoritmo Blast é o tBlastx, o qual compara as seis fases de leitura

possíveis de uma sequência, contra uma tradução das sequências do banco de nucleotídeos, feita também para as 6 fases de leitura possíveis (NCBI).

Como forte aliado na caracterização dos grupos enzimáticos tem-se o emprego da **Filogenômica**, a qual representa o conjunto de metodologias que explora as relações de parentesco entre sequências biológicas (DNA e proteínas), permitindo a predição funcional protéica com base em ortólogos/parálogos conhecidos (SJÖLANDER, 2004).

Fazem parte dessa abordagem o uso de alinhamento global e local, construção de árvores filogéticas, busca em diferentes bancos de dados por genes homólogos, “motifs”, e domínios enzimáticos conservados, empregando ferramentas de bioinformática; seguido do cruzamento dos dados com informações biológicas presentes na literatura, auxiliado pelo uso de índices de bioestatística.

1.3.4. Metabólitos secundários

Metabólitos secundários tem sido tradicionalmente caracterizados como compostos não essenciais para a sobrevivência do organismo, produzidos no final do ciclo celular (principalmente durante a fase estacionária) e são geralmente entendidos como um tipo de mecanismo de defesa contra competidores, à parte do metabolismo geral da célula, com processos de regulação e fisiologia próprios (ZHANG, FERREIRA & TANG, 2010).

Todavia, atualmente sabe-se que a produção destes compostos estão interligados aos demais constituintes do metabolismo, compartilhando reguladores e precursores metabólicos e influenciando na fisiologia geral do organismo. Um claro exemplo é o caso do metabólito secundário da classe das fenazinas, a piocianina, produzida pela espécie bacteriana *Pseudomonas aeruginosa*, a qual interfere na Rede *Quorum Sensum* (rede “QS”) de regulação gênica (ZHANG, FERREIRA & TANG, 2010).

A rede “QS” compreende uma cascata de *motifs* e propagação de sinais através de quinonas e homoserina-lactonas, e atua na regulação da expressão de vários genes,

bem com na regulação da fisiologia bacteriana em respostas a identificação de intensidades específicas de sinais bioquímicos e demais alterações ambientais. A interferência da piocianina neste mecanismo demonstra sua clara interligação ao metabolismo primário, ocupando papel-chave para a sobrevivência dos organismos, atuando em vários momentos do ciclo celular, o que desafia as definições clássicas de metabólitos “secundários” (ZHANG, FERREIRA & TANG, 2010).

1.3.5. Actinomicetes: riqueza de metabólitos secundários

Actinomicetes são micro-organismos comumente encontrados em solos e outros ambientes, conhecidos por sua habilidade de produzir uma vasta quantidade de metabólitos secundários distintos, principalmente antibióticos. Outra característica interessante dos actinomicetes, que lhes confere vantagem adaptativa em relação à muitos outros grupos de bactérias, é a capacidade de produzir esporos como formas de resistências a ambientes desfavoráveis (ZHANG, FERREIRA & TANG, 2010).

A Figura 1 ilustra a diversidade do grupo, tendo por base organismos obtidos por técnicas de isolamento e cultivo nos mais diversos ambientes. Nesta figura é possível identificar a supremacia do grupo das *Streptomyces* (57%) sobre os demais actinomicetes, num total de 53.300 estirpes isoladas e identificadas no mínimo até o nível de gênero.

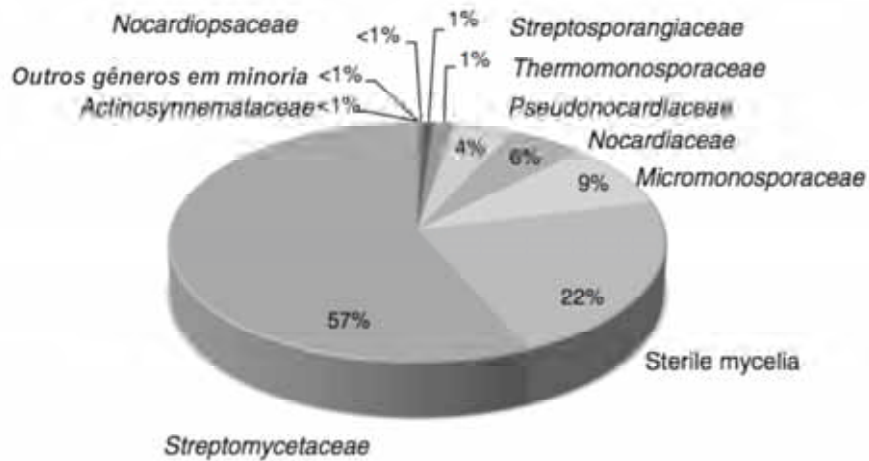


Figura 1. Representação da diversidade de *Actinomycetes* em uma coleção 53.300 estirpes, sendo 27.400 identificadas apenas até o nível de gênero. O grupo “Sterile mycelia” representa as estirpes não esporulantes devido a características taxonômicas morfológica (Fonte: de GENILLOUD et al., 2010).

O grupo dos Actinomicetes são responsáveis pela síntese de muitos bioativos explorados pela farmacologia, sobretudo com a implementação de inovações no processo clássico de prospecção, explorando sequências gênicas conservadas de enzimas envolvidas para uma rápida identificação de agrupamentos inéditos, fazendo com que mais compostos sejam descobertos a cada ano (Fig. 2).

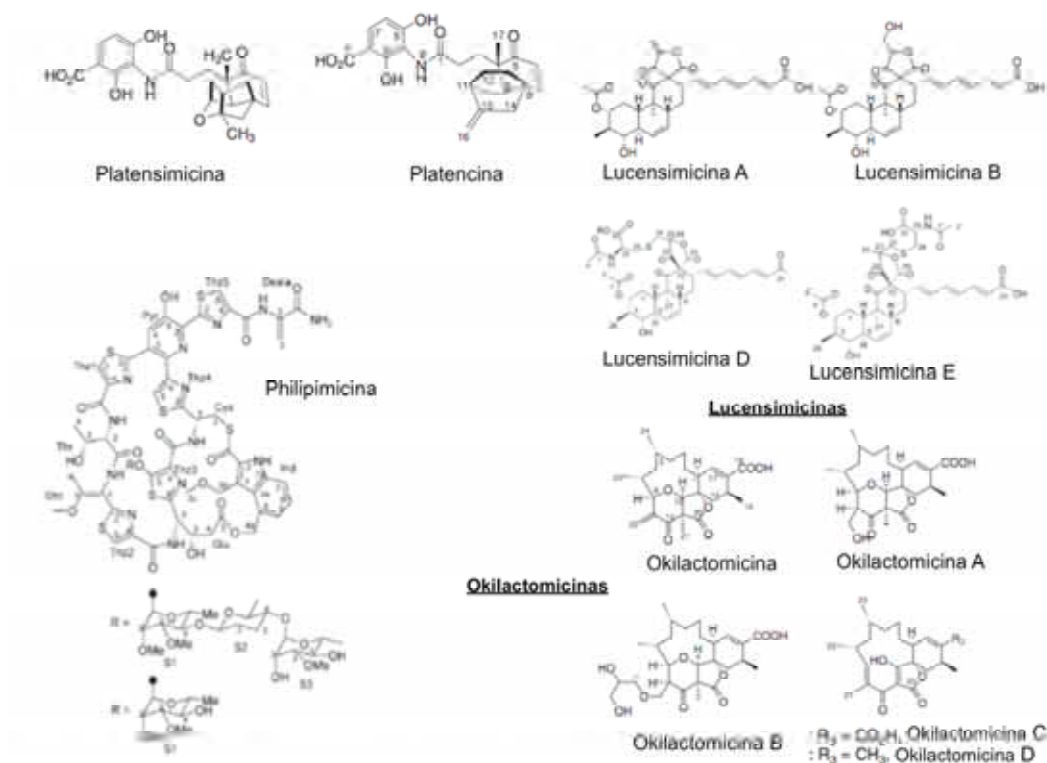


Figura 2. Alguns novos compostos produzidos pela classe *Actinomycetes* descobertos em 2010 (Fonte: de GENILLOUD et al., 2010).

Os Actinomicetes são abundantes em muitos solos. GENILLOUD e colaboradores (2010) isolaram e identificaram, por estratégia de 16S rDNA, 790 estirpes, caracterizando-as quanto a gênero e espécies. A maioria absoluta dos gêneros *Streptomyces* e *Micromonospora* inviabilizou a sua inclusão comparativa quanto a número de espécies e estirpes por gêneros (Fig. 3), indicando um vasto horizonte a ser explorado. Os dados demonstram o quanto o grupo é importante para o ambiente edáfico. De fato, em muitos ambientes o grupo bacteriano chega a ocupar nichos frequentemente pertencentes a fungos, como a degradação de compostos recalcitrantes como celulose e lignina.

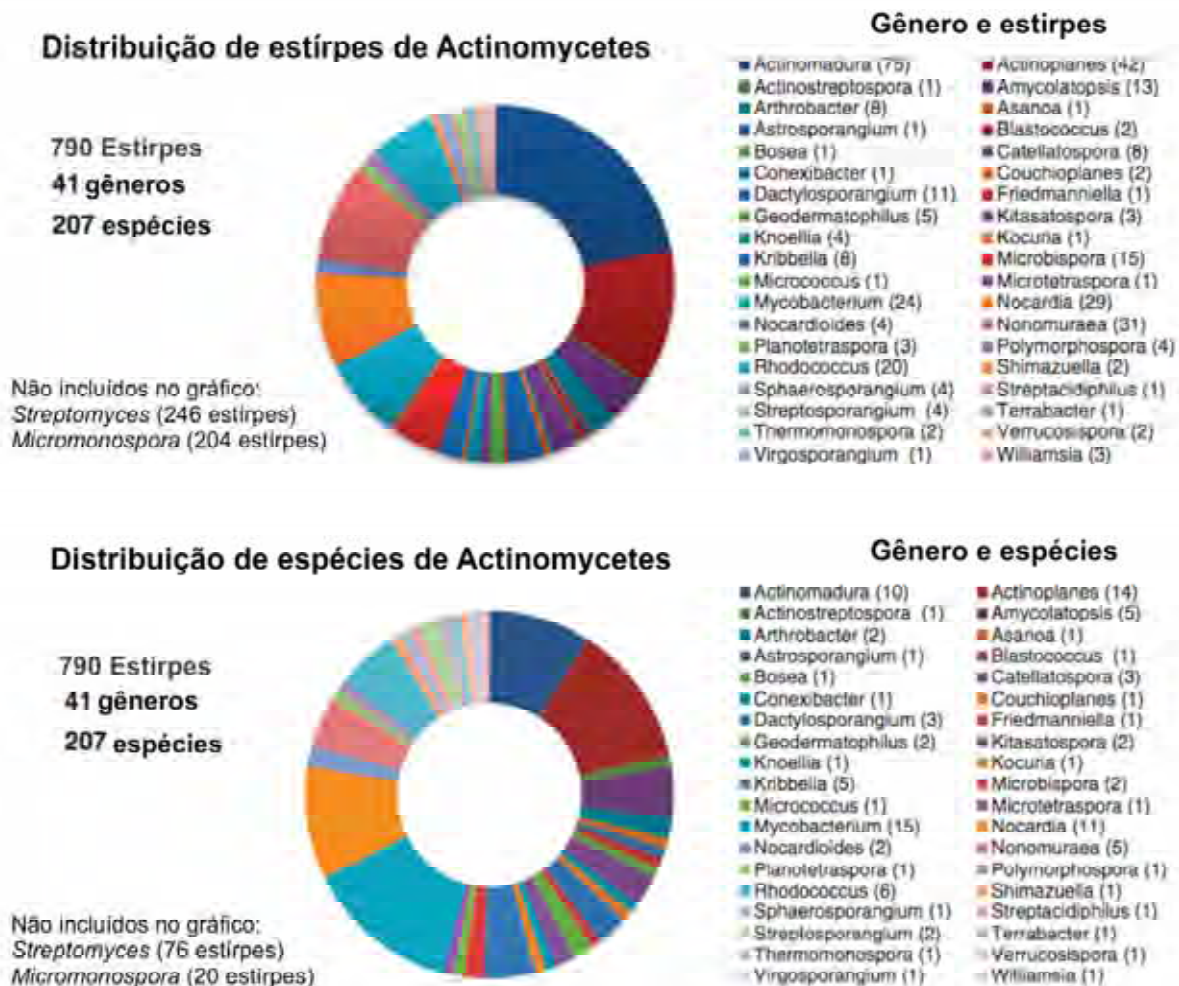


Figura 3. Representação da diversidade de Actinomycetes acessados por isolamento sistemático em amostras de solo, seguido de sequenciamento parcial de 16S rDNA (Fonte: de GENILLOUD et al., 2010).

Dentre os actinomicetes, encontram-se linhagens que abrigam em seu genoma genes que codificam para a PKS I e a PKS II, e mesmo para as duas formas simultaneamente (BUSTI *et al.*, 2006). Algumas linhagens possuem em seu genoma mais do que um agrupamento gênico para a PKS II e, frequentemente, a mesma rota que leva a produção de metabólitos secundários com atividade antimicrobiana é utilizada na síntese de pigmentos de esporo nesse grupo (HOPWOOD, 1997).

1.3.6. O gênero *Streptomyces* e a produção de metabólitos secundários

Segundo descrição feita por OMURA e colaboradores (2001), *Streptomyces* é um gênero de bactéria gram-positiva, cujo crescimento pode ocorrer em solos, pântanos, e habitats de costa marinha. Em seu desenvolvimento, estes organismos procariotos formam micélios filamentosos similares aos encontrados em fungos (eucariotos). A diferenciação morfológica em *Streptomyces* envolve a formação de uma camada de hifas aéreas na superfície da colônia que se projetam para o ar adjacente e se diferenciam em cadeias de esporos. Esse processo, único dentre as bactérias gram-positivas, requer uma coordenação especializada do metabolismo, e é mais complexo do que nas demais. Os autores ressaltam também que a propriedade mais intrigante deste gênero é a ampla capacidade de produção de metabólitos secundários, dentre os quais é notável a produção de antibióticos e compostos bioativos importantes para os segmentos da medicina humana e veterinária e para a agricultura, dentre outros fins biotecnológicos.

Assim, *Streptomyces* é reconhecida como uma rica fonte de metabólitos secundários, nos quais intermediários comuns do metabolismo básico celular são condensados para a formação de uma miríade de moléculas de estruturas complexas com ligações químicas e combinações não usuais para as demais vias do metabolismo primário, através de rotas bioquímicas peculiares, muitas vezes diferenciadas para micro-organismos dentro de uma mesma espécie. Outra particularidade é a ocorrência em uma mesma colônia de uma produção não de apenas um composto, mas de uma mistura de compostos diferentes dentro de uma mesma família química, dentre os quais o composto predominante é influenciado pelas características físico-químicas do meio e o tipo de elemento precursor disponível no momento (MARTIN & DEMAINE, 1980). Sendo assim, alguns pesquisadores empreendem esforços na definição de combinações específicas de certas condições de cultura para estimular a produção de uma molécula de interesse a qual possa ser de produção pouco expressiva nas condições de ambiente natural ou padrões de cultivo em laboratório. Contudo, a abordagem mais promissora para síntese de novos compostos bioativos está

relacionada ao uso da Metagenômica para busca de novos genes biossintéticos para clonagem e expressão em hospedeiros heterólogos, ou mesmo um hospedeiro heterólogo nativo (do mesmo gênero) que não detenha previamente esta referida via.

Consequentemente, para o material gênico de uma nova enzima ou grupo enzimático recuperado por metagenômica realmente ser expresso é necessário a sua introdução em um organismo vivo capaz de executar as instruções da informação gênica resultando na tradução da enzima (ou conjunto destas) na conformação ativa. Isso significa que o sucesso da pesquisa que vise produção de novas moléculas por esta estratégia é condicionado à disponibilidade de um hospedeiro heterólogo compatível com a expressão correta do gene exógeno, ou mesmo hospedeiros nativos com vias adaptadas para não interferir na detecção dos novos produtos. No caso de genes relacionados à PKS II do grupo das *Streptomyces* já estão disponíveis como hospedeiros heterólogos nativos as linhagens de *Streptomyces lividans*, *S. albus* e *S. coelicolor*, e como hospedeiros heterólogos foram desenvolvidas linhagens mutantes de *Escherichia coli* desenvolvidas especificamente para expressão de genes do grupo (MARTIN & DEMAIN, 1980; WANG, 2000; BALTZ, 2006).

A disponibilidade de tais linhagens é um ponto-chave na busca por novos fármacos, uma vez que particularidades da sequência gênica das *Streptomyces* reduzem significativamente a compatibilidade com outros possíveis hospedeiros normalmente empregues para a expressão heteróloga de outros grupos.

Um dos fatores mais limitantes para o grupo em questão para a expressão de proteínas exógenas tem relação com o próprio conjunto de códons preferenciais (“*Códon Usage*”) presentes na maior parte do genoma de *Streptomyces*. O códon de uso pode atuar como fator limitante para o intercâmbio entre o hospedeiro nativo do gene de interesse e o hospedeiro heterólogo que será empregue na estratégia de expressão. Isso ocorre porque se um conjunto de trincas predomina em um genoma, é natural que a maioria dos tRNAs (RNAs transportadores) disponíveis no organismo corresponda a este códon preferencial. Assim a tradução de uma proteína exógena cujo códon preferencial apresente baixa ocorrência no candidato a hospedeiro não ocorrerá de forma correta, ou apresentará um baixo nível de expressão em decorrência da

carência de códons (tRNAs) correspondentes às trincas raras para o genoma do hospedeiro heterólogo, sem os quais a adição do aminoácido referente àquela porção do transcrito não ocorre.

Utilizando um mapa bi-dimensional da distribuição dos códons de uso preferenciais dos organismos vivos mais estudados no meio acadêmico, GUSTAFSSON e colaboradores (2004) teceram considerações sobre os grupos de organismos que mais compartilhavam códons preferenciais entre si, sendo que estes grupos podem ser deduzidos com base na proximidade espacial destes em um gráfico de estatística multivariada (Fig. 4).

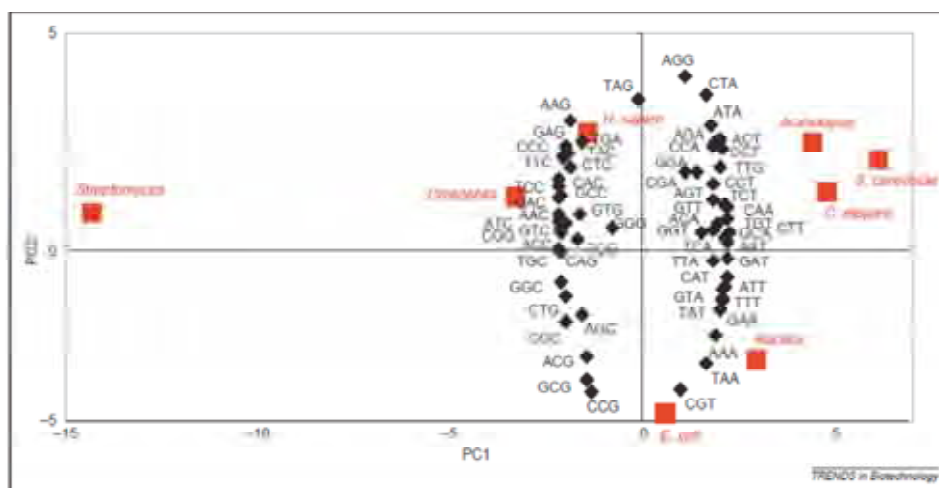


Figura 4. Representação gráfica espacial do códon de uso de organismos frequentemente estudados no ambiente acadêmico, correlacionando o grau de compatibilidade dos indivíduos quanto ao códon preferencial. Imagem adaptada do trabalho científico de GUSTAFSSON et al., Fig. 1, p.347, 2004.

Assim, com base no gráfico pode-se inferir que *Saccharomyces cerevisiae* seria o candidato mais adequado para a expressão de genes do organismo vegetal *Arabidopsis* bem como para genes de *Caenorhabditis elegans* (animal: verme do grupo dos nematódeos). Ao mesmo tempo tem-se uma proximidade maior entre o grupo das Drosófilas e a os genes da espécie humana, bem como se pode notar uma proximidade entre os gêneros bacterianos da *Escherichia coli* e o *Bacillus*. No entanto todos os sete organismos citados compartilham mais códons de uso entre si do que cada um

compartilharia com o grupo das *Streptomyces* (distante de todos os demais na conformação espacial do mapa).

O perfil de códons de uso deste último grupo foi definido pelos autores como o mais extremo dos estudados, sendo conhecido para a espécie *Streptomyces coelicolor* uma situação em que a maior parte dos códons possuem um C ou um G como nucleotídeo terminal, o que caracterizaria o alto conteúdo destas bases nitrogenadas (por volta de 71%) na espécie analisada. Este perfil de códons de uso para este gênero se constitui em um elemento explicativo para a dificuldade de expressão de genes derivados de *Streptomyces* utilizando organismo de outros gêneros como hospedeiros (GUSTAFSSON et al., 2004).

No entanto, caso o pesquisador opte pelo uso de outros organismos para a expressão, as condições de eficiência para a tradução da proteína de interesse podem ser favorecidas por meio de Engenharia Molecular. No caso específico de expressão de enzimas envolvidas na biossíntese de policetídeos do tipo II, são relatados na literatura uso de *Escherichia coli* por duas abordagens diferentes.

Em uma delas, a estratégia consiste na modificação dos códons presentes na sequência gênica para o conjunto de trincas otimizadas para as mais abundantes em *E. coli*, e portanto para as quais a maquinaria do hospedeiro heterólogo é plenamente funcional (BALTZ, 2006). A outra abordagem, por sua vez, consiste em modificações ao longo do genoma do próprio hospedeiro heterólogo (*Escherichia coli*) com o objetivo de aumentar a ocorrência de códons raros limitantes para a expressão de genes derivados de *Streptomyces*. A vantagem de tais linhagens mutantes de *E. coli* é a versatilidade de uso de um mesmo organismo para a expressão de genes policetídeos distintos, aliado ao crescimento rápido da *E. coli* em comparação com a *Streptomyces*. No entanto, mesmo que estes mutantes se mostraram aptos na expressão enzimática e produção de estruturas simples de policetídeos, o rendimento destes metabólitos é maior para hospedeiros heterólogos do grupo das *Streptomyces* (BALTZ, 2006; PEIRU et al. 2008).

1.3.7. Análise do setor farmacológico e perfil das organizações envolvidas na descoberta de novas drogas

Desde a descoberta da estreptomicina na década de 50, produzida por um actinomicete isolado do solo, microbiologistas e farmacêuticos do mundo inteiro passaram a investir grandes esforços no cultivo de diversos micro-organismos para pesquisar novas linhagens capazes de sintetizar antibióticos. A busca por novos compostos levou as companhias farmacêuticas a isolarem dezenas de milhares de linhagens produtoras, e ainda hoje micro-organismos cultiváveis de solo representam a principal fonte de antibióticos e outros compostos ativos. Entretanto, os métodos tradicionais para pesquisa de novas drogas que envolvem o cultivo de micro-organismos isolados do solo não são mais tão promissores, devido principalmente à alta taxa de redescoberta de antibióticos já conhecidos, que chega a 99,9% (ZAEHNER & FIELDLER, 1995). Durante um grande período de tempo houve uma refreada na descoberta de novas moléculas, chegando ao limite da metodologia clássica baseada somente em isolamento, cultivo e bioensaios. Esse “Gap” de inovação se estendeu desde o início da década de 1970, estagnando na descoberta das quinolonas e estreptograminas e foi superado no ano 2000, com a descoberta de uma nova classe: oxazolidinona. Entre 2000 e 2009, foi ainda descoberta mais uma classe, a dos lipopeptídeos, além de novas estruturas para a classe das mutilinas (FISCHBACH & WALSH, 2009).

Conseqüentemente, nos últimos 30 anos, apenas duas classes de antibióticos inéditas foram lançadas no mercado: a oxazolidinona linezolida e o lipopeptídeo cíclico daptomicina. Como agravante, a telavancina (um glicopeptídeo), a retapamulina (uma pleuromutilina), e a tigeciclina (última geração de tetraciclinas) foram os únicos princípios ativos de antibióticos admitidos como moléculas totalmente novas pela “US Food and Drug Administration” (FDA) lançados no mercado nos últimos 5 anos (HAMAD, 2010). Em adição a estas, são poucos os antibióticos em circulação ainda considerados seguros quanto a retenção da capacidade ativa contra micropatógenos multi-droga resistentes (HAMAD, 2010).

Esse declínio é também marcado por um caráter desigual da distribuição de moléculas dentro das diferentes classes de antibióticos, marcadamente elitista com a predominância de compostos policetídeos, que chegaram a perfazer mais de um terço do total das drogas lançadas no mercado no período de 2005 a 2007 (WEISSMAN, 2009). A Figura 5.A ilustra os rendimentos obtidos em bilhões de dólares para cada um dos principais princípios ativos dos fármacos mundialmente comercializados referentes ao ano de 2009, com destaque para os policetídeos da classe das cefalosporinas, responsáveis por 11,9 bilhões de dólares. Contudo o cenário de crise do setor transparece na morosidade do caminho percorrido entre a prospecção e a chegada ao mercado consumidor. A Figura 5.B ilustra o perfil da distribuição de moléculas em fase de pré-implantação em 2010, e demonstra que mais do que o próprio processo de descoberta de novos bioativos, o grande gargalo do processo é representado pelo período pré-clínico (Fig. 5.B, HAMAD, 2010).

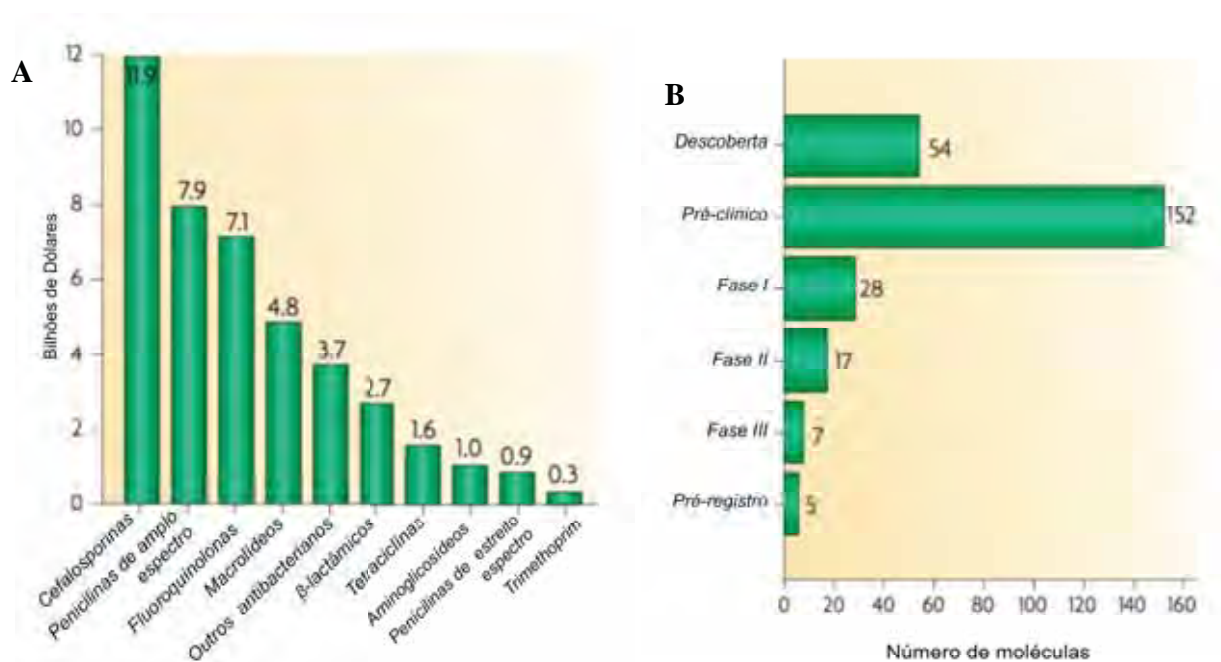


Figura 5. Distribuição por família química e por estágio de pesquisas para novas drogas (HAMAD 2010). A: Faturamento em bilhões de dólares para os antibacterianos mais utilizados. B: "Pipeline" da produção de novos fármacos.

Outro fator que contribui muito para o tempo de retenção de cada droga em cada estágio é a natureza de sua aplicação e a urgência na demanda de certos grupos clínicos (Fig. 6, KAITIN, 2010).

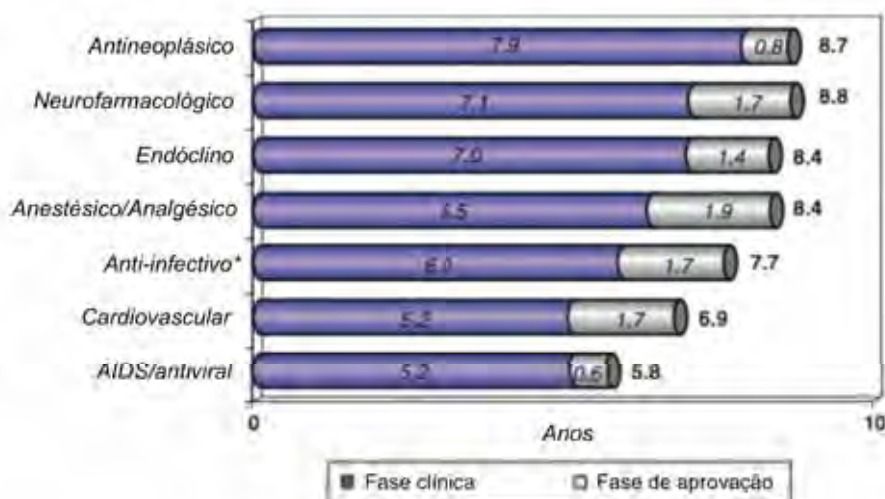


Figura 6. Médias obtidas para o tempo de retenção de novas moléculas de acordo com a área patológica, obtidos no período de 2003 à 2007. * A categoria Anti-infeccioso exclui agentes antivirais para a AIDS (detentores de categoria própria) (KAITIN, 2010).

Como pode ser observado na Figura 6, os medicamentos para o combate ao HIV são os mais rapidamente implementados, devido em parte a rapidez em que as drogas utilizadas atingem a estágio de ineficácia, frente a novos sorotipos prevalentes da doença. Tanto essa classe de medicamentos quanto os antineoplásicos, apresentam também o menor tempo necessário a regulamentação (Fase de aprovação), refletindo uma clara redução do processo burocrático envolvendo a liberação destes fármacos, garantido um intercâmbio mais dinâmico entre o setor acadêmico/técnico/científico e a comunidade.

Outro aspecto a ser levado em conta é como se dá o processo de geração de novas drogas, identificando quais são as instituições que mais contribuem mundialmente para a sua geração. A Figura 7, por exemplo, ilustra a origem das 252 novas drogas aprovadas pela “US Food and Drug Administration” no período de 1998 a 2007.

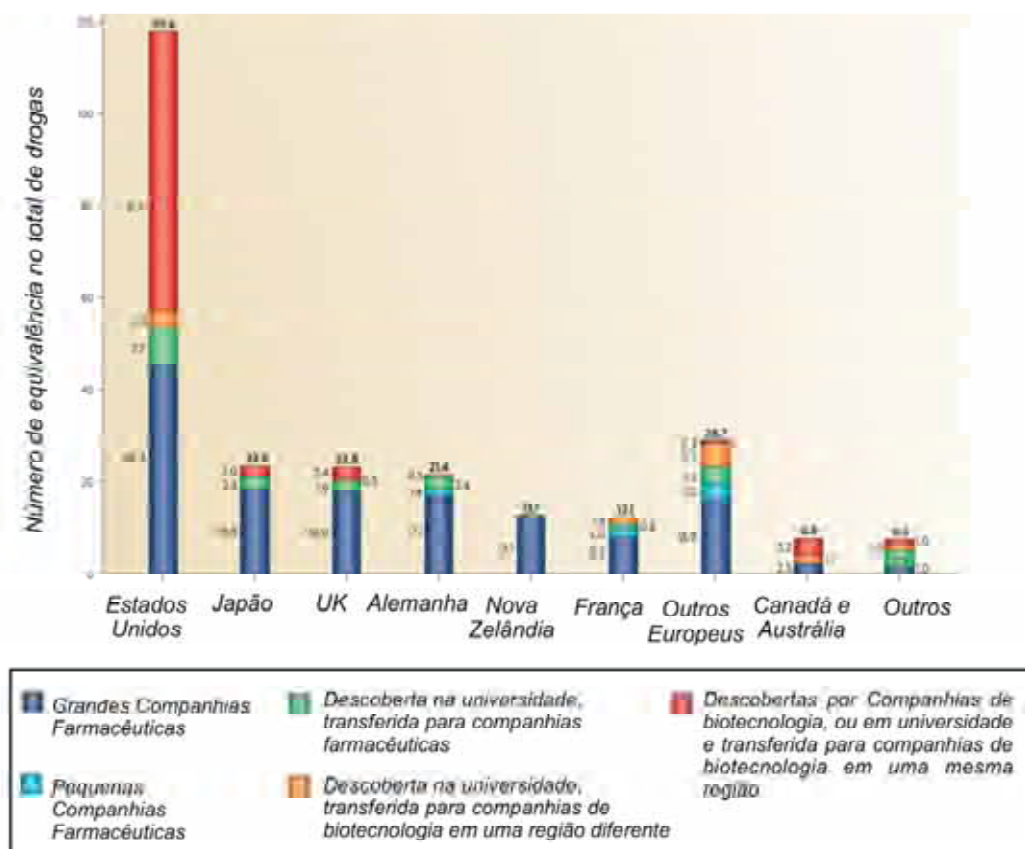


Figura 7. Contribuições da iniciativa pública e privada para a descoberta de novos fármacos em diferentes países, para o período entre 1998 e 2007 (KNELLER, 2010).

Como pode ser observado no gráfico, os EUA são responsáveis pela maior parte das drogas geradas, e a universidade representou um importante foco de geração de novos produtos para o setor privado em biotecnologia, superando o rendimento de grandes companhias farmacêuticas (KNELLER, 2010). Estes dados não deixam dúvidas sobre o papel das universidades na geração de novos medicamentos, claramente demonstrando como o setor público e privado poderiam investir para alavancar a produção de novos fármacos.

No Brasil, a política de regulamentação de novas drogas quanto a prioridades de implementação, é responsabilidade da Política Nacional de Medicamentos (PNM) e da Agência Nacional de Vigilância Sanitária (ANVISA). A PNM foi criada em 1998, e foi responsabilizada por regulamentar a lista das drogas consideradas essenciais e pelas

normativas de segurança para os medicamentos, recursos humanos, e desenvolvimento técnico científico. A lista de drogas essenciais já existia, mas não fora previamente revista por cerca de 16 anos. A PNM foi responsável por reformular a lista em 1999, e pelas atualizações da mesma em 2002 e 2006. A ANVISA, por sua vez, foi criada em 1999 com prioridade para responder sobre questões da vigilância sanitária (VIDOTTI; DE CASTRO & CALIL, 2008).

Um passo importante para a saúde brasileira, onde 70% da população usufrui de estabelecimentos públicos de saúde, foi a aprovação da política de drogas genéricas, em 1999, com a quebra de patentes. Quatro anos depois da aprovação, os medicamentos genéricos já ocupavam 12% do total comercializado no país. Do restante, 68% correspondiam a medicamentos similares (não – patenteados), e 20% correspondiam aos medicamentos protegidos por patentes (VIDOTTI; DE CASTRO & CALIL, 2008). Tal prática permitiu um acesso mais democrático às drogas consideradas essenciais para a saúde.

A importância dos genéricos e similares é inegável uma vez que, realmente, representam opções mais atrativas em um setor em que é comum a prática da mesma formulação diferir unicamente em marcas e preços. WADT (2003) analisou os novos medicamentos lançados no mercado no período de 1998 à 2001 (Fig. 8). Dos 285 medicamento “novos” considerados pela ANVISA, 84 foram considerados não inovadores (vermelho), por na verdade apresentar as mesmas formulações de outros medicamentos previamente lançados no mercado. Do restante, 47 não eram alvo do estudo do autor (amarelo), e 154 fármacos (em azul) realmente continham uma nova molécula química em sua formulação.

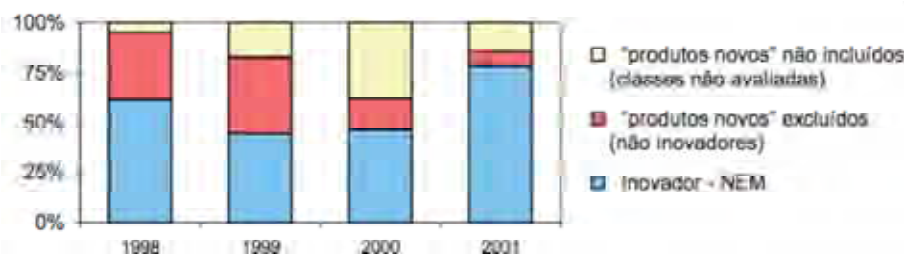


Figura 8. Avaliação de novos medicamentos lançados no Brasil no período de 1998 à 2001 (Fonte: WADT, 2003).

Historicamente, as corporações farmacêutica que atuavam no Brasil tanto com capital nacional, quanto transnacional, tem se voltado para a manipulação de produtos farmacêuticos, relegando as iniciativas para a descoberta de novas formulações à um segundo plano, com recursos escassos. Felizmente essa realidade está começando a se modificar, com o aumento dos investimentos e o envolvimento das companhias em todos os estágios para o desenvolvimento e produção de novas drogas, o que parece ter uma estreita relação com as mudanças nas políticas de normativas para o setor, a ampliação da capacidade de construção, no ambiente econômico favorável, aliados a consolidação da pesquisa nas universidades. De fato, um aumento na criação de grupos de estudos e uma maior interação entre estes e as companhias farmacêuticas tem ampliado as capacidade de pesquisas voltadas para novos produtos, culminando na primeira droga desenvolvida inteiramente no Brasil no ano de 2004: o anti-inflamatório alpha-humuleno (VIDOTTI; DE CASTRO & CALIL, 2008), derivado da espécie vegetal *Cordia verbenácea*, conhecida pelo nome popular de “erva-baleeira”.

Tendo em vista o exposto, percebe-se que o mercado mundial de medicamentos ainda é dominado por grandes Companhias Farmacêuticas e de Biotecnologia, com participação ativa do campo acadêmico no desenvolvimento de novas formulações, e que os Estados Unidos é o maior produtor de medicamentos. Outros países desenvolvidos rateiam entre si as responsabilidades pelo restante das novas formulações. Embora ainda em seus primeiros passos, o Brasil tem apresentado avanços no desenvolvimento técnico/científico de novos medicamentos, contudo ainda há muito o que ser feito para um acesso democrático aos medicamentos, e investimentos em pesquisa acadêmica parece ser um caminho promissor, já triado por países desenvolvidos em implementações de estratégias bem-sucedidas de parcerias entre a Universidade e a Iniciativa Público/Privada.

1.3.8. Prospecção de antibióticos empregando Biologia Molecular

Para a descoberta de novos genes envolvidos na biossíntese de antibióticos,

uma abordagem bastante promissora é a utilização de oligonucleotídeos iniciadores degenerados para amplificar sequências homólogas em uma biblioteca metagenômica. Estes oligonucleotídeos são sintetizados para hibridização com sequências conservadas dos agrupamentos gênicos (METSÄ-KETELÄ *et al.*, 1999; COURTOIS *et al.*, 2003, SCHUCH, 2007; GOMES, 2008; MASSINI, 2009).

Essa estratégia de prospecção é bastante utilizada no caso da prospecção por antibióticos policetídeos, de forma que fragmentos de DNA ambiental clonados são utilizados como molde para amplificação da região conservada de genes de PKS (COURTOIS *et al.*, 2003).

Como em muitos casos os genes da PKS encontram-se agrupados em um mesmo local no cromossomo (HOPWOOD, 1997), pode-se encontrar um agrupamento gênico inteiro, adjacente à sequência amplificada por PCR. Para tanto, o clone de amplificação positiva na PCR é submetido à subclonagem e sequenciamento, seguido de análise de homologia com sequências nucleotídicas depositadas em banco de dados como o Genbank (NCBI).

Metsä-Ketelä e colaboradores (1999) desenvolveram um método eficiente para a prospecção de genes pertencentes ao grupo da *pks mínima*. Os autores desenharam oligonucleotídeos iniciadores degenerados capazes de amplificar fragmentos de genes de cetosintases a partir de diversas linhagens de *Streptomyces spp.* produtoras de antibióticos policetídeos e pigmentos, e também de bactérias isoladas de solo. Utilizando as sequências de aminoácidos obtidas a partir dos produtos amplificação foi construída uma árvore filogenética para a região KS α da PKS II. Sequências conhecidas para linhagens de *Streptomyces* obtidas no Genbank também foram utilizadas na construção da árvore para ampliar as análises filogenéticas. Conforme pode ser observado na Figura 9, o resultado desse agrupamento foi a separação das sequências obtidas em dois ramos, um correspondente à classes de policetídeo-sintases envolvidas na produção de pigmentos de esporo bacteriano, e outro relacionado à produção de antibióticos policetídeos. Também foram discriminados subgrupos dentro do ramo para antibióticos.

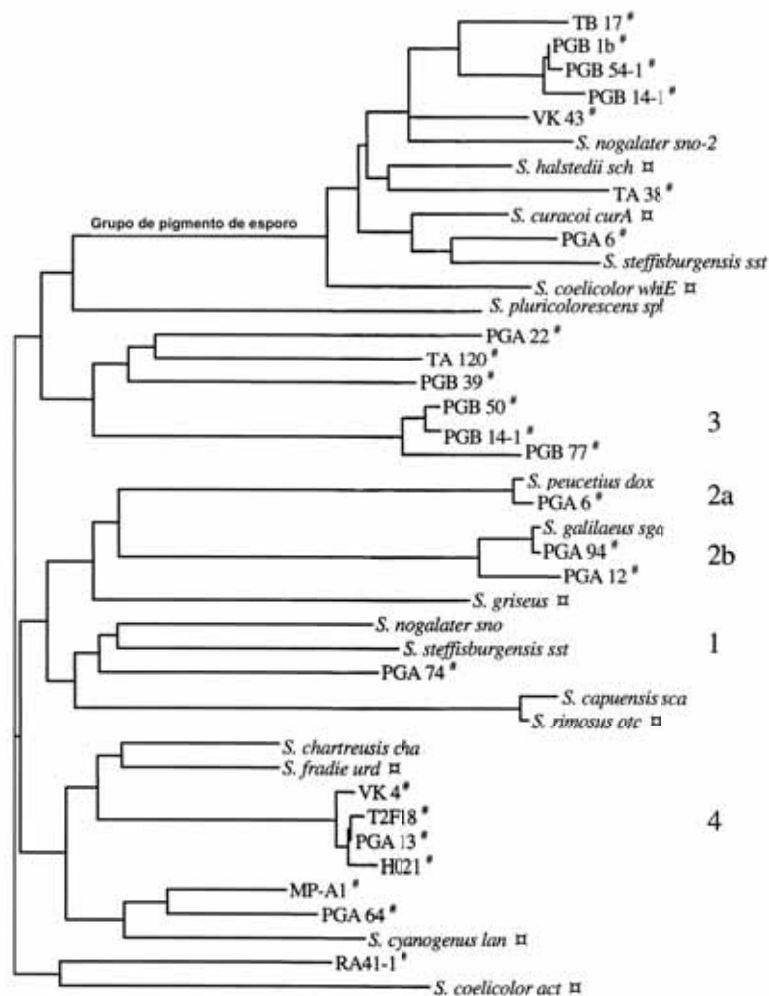


Figura 9. Árvore filogenética construída de seqüências de aminoácidos de regiões amplificadas por PCR genes *KSα* de Actinomicetes. A árvore foi construída com o programa ALIGN X (VECTOR NTI SUITE software package) utilizando configurações padrões. Seqüências marcadas com o símbolo: □, são seqüências obtidas junto ao GenBank e aquelas marcadas com # são isolados bacterianos não identificados de uma coleção feita por Galilaeus Oy. 1: antraciclinas contendo 10 acetatos de extensão, podendo iniciar com malonamida (*S. capuensis* e *S. rimosus*) ou propionato (demais); 2a e 2b: antraciclinas iniciadas com propionato mais 9 acetatos. 3 e 4: grupos com vias ainda não elucidadas, o isolado PGB39 (3) apresentou a produção de uma aglicana diferenciada (METSÄ-KETELÄ et al., 1999).

Neste trabalho os autores desenharam vários “primers” e realizaram testes de amplificação por PCR utilizando diversas linhagens de *Streptomyces* produtoras de policetídeos aromáticos e 29 actinomicetos não identificados isolados de amostras de solo. A análise dos “amplicons” por sequenciamento permitiu identificar o par de oligonucleotídeos iniciadores mais indicado para a amplificação dessa região (METSÄ-

KETELÄ *et al.*, 1999).

Esse par de iniciadores desenhado pelos autores já foi empregue em muitos outros trabalhos de prospecção de genes de policetídeos aromáticos, incluindo trabalhos envolvendo estratégias em metagenômica, o que comprova sua eficácia para recuperar genes de PKS de *Streptomyces* e outros genes, a despeito do desenho dos oligonucleotídeos obedecendo o singular códon de uso das estreptomicinas, indicando também as vantagens do uso de iniciadores degenerados para aumentar as possibilidades de recuperação de uma dada sequência de interesse (BUSTI *et al.* 2006, GOMES 2008; MASSINI, 2009).

1.3.9. Policetídeos

Os compostos policetídeos assumem lugar de destaque entre as drogas de origem natural conhecidas. Mais de um terço dos produtos naturais e seus derivados aprovados como drogas entre 2005 e 2007 foram moléculas deste tipo químico (BUTLER *et al.* 2008), sendo que os rendimentos anual de produtos farmacêuticos frequentemente fica acima de US\$ 20 bilhões. Dentre os medicamentos de maior destaque, podem-se citar a estatina, avaliada como o cardiovascular mais efetivo; e o enedieno, considerado o agente anti-carcinogênico mais poderoso já descoberto, devido a sua alta reatividade (WEISSMAN, 2009).

A riqueza das estruturas de policetídeos deriva das modificações pós-síntese das cadeias policetílicas, responsáveis na maioria das vezes pela própria reatividade da molécula. A principal forma de modificação é por oxidação, e em seguida vem as reações de transferência de grupos químicos, tais como as glicosilações e acilações. A diversidade de moléculas também abrange o tamanho diferenciado das cadeias de policetídeos (Fig. 10), indo das menores moléculas com seis a oito carbonos, como o ácido triacético lactona (TAL) e o ácido 6-metil salicílico (6-MSA), até as gigantescas moléculas da maitoxina, que possui um peso molecular de 3.422 Daltons, e uma cadeia que incorpora 164 átomos de carbono (WEISSMAN, 2009).

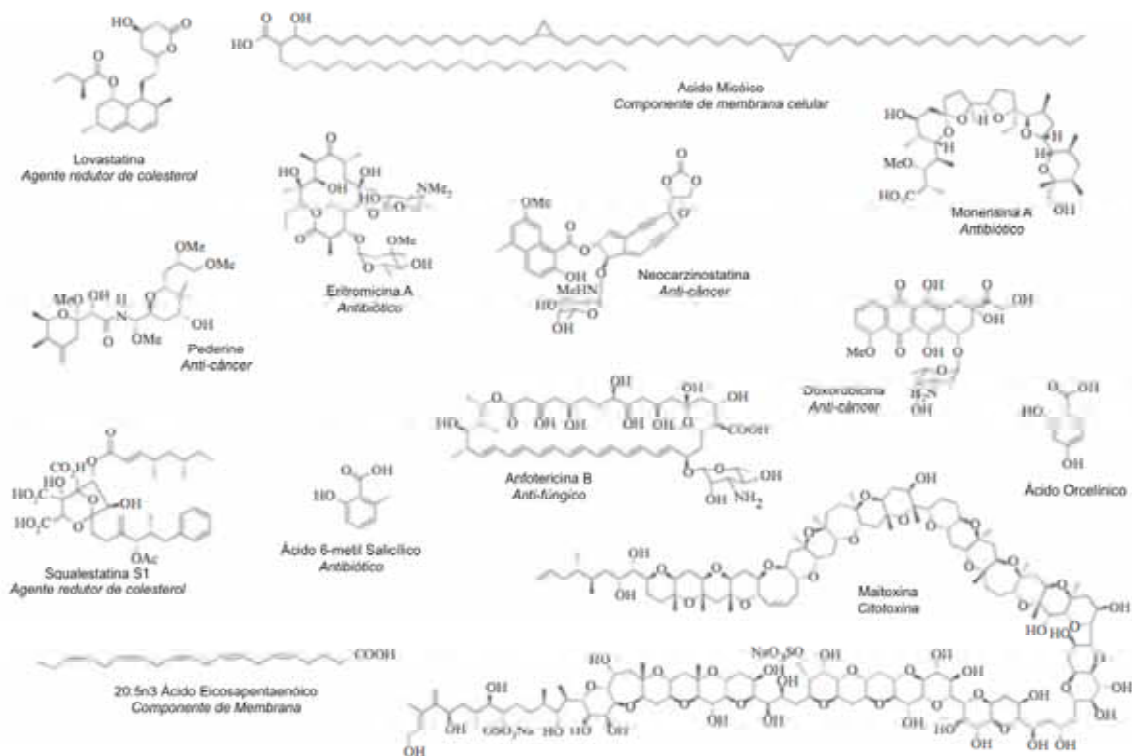


Figura 10. Estrutura de policetídeos representativos, incluindo bioativos proeminentes. (Fonte: WEISSMAN, 2009).

Os policetídeos são originalmente produtos do metabolismo secundário de algumas espécies de bactérias, fungos, plantas, e mesmo microalgas como dinoflagelados, e compõem a estrutura química de muitos fármacos: antibacterianos, antifúngicos, anticarcinomas, imunossupressores, agentes redutores de colesterol, analgésicos, neurotoxinas, antiparasitas e determinantes de virulência (RIX et al., 2002). Eles são sintetizados por complexos enzimáticos variados, denominados Policetideos Sintases (PKSs).

1.3.10. Tipos de PKS e Enzimologia dos Grupos

O complexo de enzimas que estão relacionadas à produção de policetídeos é dividido em três grandes famílias homologas às “Fatty Acid Synthases” (FAS): as “Polyketide Synthases” (PKSs) do tipo I, II e III. Tal divisão não é rígida ou definitiva, podendo haver subfamílias e “clusters” híbridos (SHEN, 2003; WEISSMAN, 2009; JENKE-KODAMA & DITTMANN, 2009). Dentre os actinomicetos encontram-se linhagens que abrigam simultaneamente em seu genoma genes que codificam para a PKS I e a PKS II (BUSTI *et al.*, 2006).

Como características distintivas, tem-se que as PKSs do tipo I (PKS Modular/Interativas) são enzimas grandes, compostas por multi-domínios, que comportam uma série de sítios para as etapas de síntese de policetídeos. Exemplos de produtos naturais derivados dessas enzimas multi-catalíticas são as classes dos macrolídeos (eritromicina) ou polienos (nistatina). As PKSs do tipo II (PKS aromática) são compostas por enzimas separadas mono ou bi-funcionais, as quais agem de maneira interativa durante a síntese para formar uma estrutura policetídic, que então é convertida enzimaticamente na forma cíclica para formar compostos aromáticos policíclicos, como a tetraciclina ou a doxorubicina (HUTCHINSON & FUJJI, 1995).

Já as PKSs tipo III, também chamadas de Chalconas Sintases, são proteínas relativamente pequenas, que estão principalmente envolvidas na produção de compostos importantes para as plantas, tais como flavonóides, estilbenos (composto fenólico), mas também são encontradas em bactérias e fungos. Ao contrário das outras PKSs, os intermediários policetídeos da PKS III não são ligados enzimaticamente, mas por tioésteres de coenzima - A livres (BRACHMANN *et al.*, 2007).

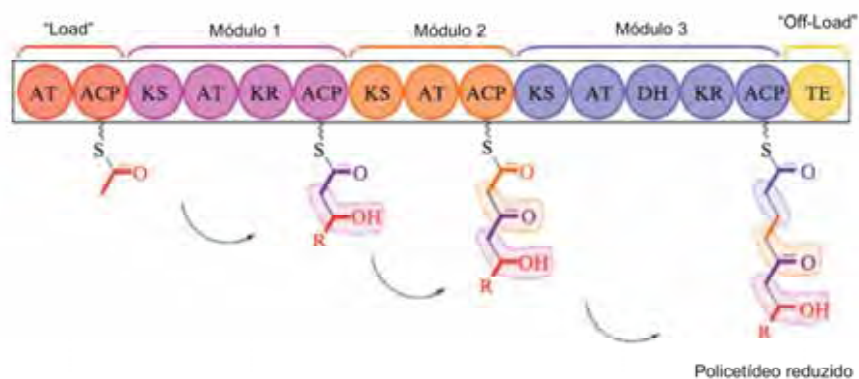
Dentre as PKSs, a via mais estudada para biossíntese combinatória são as PKS tipo I, o que possibilitou o desenvolvimento de banco de dados específicos para essa família, e um sistema eficiente de comparação de domínios enzimáticos capaz de predição da maioria dos substratos por meio de softwares baseados em homologia (YADAV, GOKHALE & MOHANTY, 2003). Tal nível de conhecimento ainda não foi atingido para o grupo de PKS II, e em virtude da sua importância ecológica e

biotecnológica é imprescindível que metodologias similares de predição sejam desenvolvidas para auxiliar o direcionamento de pesquisas em proteômica e engenharia molecular de forma mais rápida.

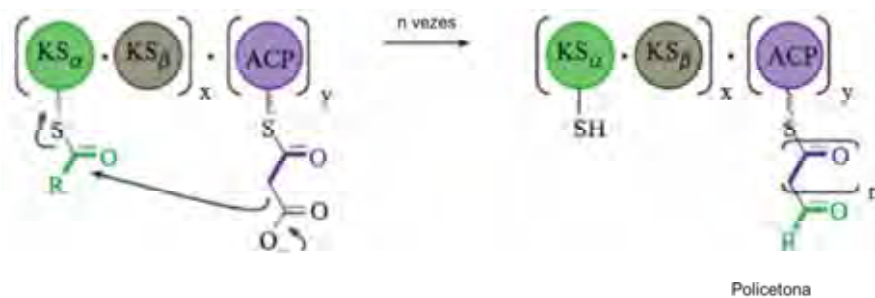
Um dos grandes pesquisadores envolvidos no estudo de PKS foi o Professor Doutor Charles Richard Hutchinson, que dedicou parte da sua vida a um estudo minucioso sobre essas importantes classes de enzimas, o que foi reconhecido recentemente por duas matérias especiais sobre a vida e o trabalho do autor na revista Nature, no periodico "The Journal of Antibiotics" (OLANO, 2011; The Journal of Antibiotics (editorial), 2011). A morte do pesquisador em 5 de Janeiro de 2010, sem dúvida alguma representou uma grande perda para o meio acadêmico. O legado de Hutchinson (OLANO, 2011), fruto de 40 anos dedicados ao estudo das PKSs envolvendo a busca por novas drogas, especialmente contra o câncer (por ironia, a doença que o vitimou aos 66 anos) compreende o estudo de policetídeos de origem bacteriana e fúngica. Dentre as PKS I, Hutchinson e colaboradores estudaram as vias de síntese da eritromicina, geldamicina, herbimicina, lasolocida, megalomicina, midecamicina, rapamicina e rifamicina. Já na família de PKS II, estudaram as vias da eloramicina, da doxorubicina, da daunorubicina, da fredericamicina, jadomicina e tetracenomicina. Sem dúvida, sua contribuição acadêmica e profissional representou um marco para a pesquisa de novos medicamentos.

A Figura 11 apresenta esquemas de mecanismo de reação para os três tipos de complexos enzimáticos, ilustrando a dinâmica geral da forma de atuação entre os grupos catalíticos.

A PKS I (Modular / Não Interativa)



B PKS II (Aromática / Interativa)



C PKS III (Interativa e Independente de ACP)



Figura 11. Esquema de reações e dinâmica catalítica das PKSs (WEISSMAN, 2009).

Como pode ser observado na Figura 11, a ação das PKSs Modulares (Fig. 11.A) é do tipo não interativo. A reação se inicia no módulo de ativação, o "loading module", pela ligação tioéster do precursor inicial ao ACP por ação do primeiro sítio AT. Existe

uma transferência do substrato ativado para o ACP do próximo módulo, em uma reação conhecida como tipo “ping-pong” entre sítios ativos, e as enzimas deste módulo irão atuar para extensão de um passo da reação para adição de um novo monômero à cadeia cetônica, e para modificação desse grupo por ações óxido-redutoras e de transferases. Essa nova molécula será por conseguinte o substrato do próximo módulo. O passo final da reação é a ação da Tio-esterase (TE) que irá liberar a moléculas de policetideo da ACP do módulo de terminação: “Off Module” (WEISSMAN, 2009). Vale lembrar que esse não é o único tipo de reação de grupos de PKS I: Existem as PKS I *cis* AT e *trans* AT, e as PKS I interativas (um ou dois módulos revesam entre si os passos de reação, ou o mesmo gene dá origem a várias cópias do polipeptídeo, que irá atuar de forma interativa nas reações).

Por conseguinte, tanto a PKS II (Fig. 11.B) quanto a PKS III (Fig. 11.C) são interativas. Contudo, a ação das PKS II na síntese da cadeia cetônica inicial envolve a atuação interativa de três subunidades enzimáticas: KS α , KS β e ACP. Primeiro ocorre uma ligação tioéster do precursor inicial (maioria Acetil-CoA) ao ACP, reação catalisada pela KS α , mas que pode também ocorrer de forma espontânea *in vitro*, depois a primeira unidade da cadeia é construída por uma reação de condensação tipo “Claisen” (denominação da reação entre dois ésteres, ou um éster e um grupo cetona, na presença de uma base forte ou um sítio catalítico, produzindo um β -ceto-éster ou β -dicetona, respectivamente), de um novo substrato (Malonil-CoA e Metil-malonil-CoA), catalizada pelas subunidades KS α e KS β , sobre o início da cadeia ainda ancorada na ACP. A reação tipo “Claisen” se repete muitas vezes, com a cadeia nascente sempre ancorada à ACP, e com atuação constante das mesmas enzimas, o que lhe confere o título de interativa. Durante a síntese, unidades enzimáticas adicionais atuam sobre essa cadeia, produzindo grupos cíclicos e aromáticos, ou reduzindo os grupos cetônicos. Após a liberação da molécula da ACP, muitas outras enzimas atuam sobre a cadeia em modificações pós sintéticas variadas, como a ação de oxidases, transferases e hidroxilases (WEISSMAN, 2009).

A PKS III, por sua vez, é surpreendentemente constituída apenas por uma unidade KS, e o precursor inicial é derivado por ligação direta da acetil-CoA livre ao sítio

catalítico da KS, que realiza sozinha todos os passos da ligação tipo “Claisen”. O potencial redox do sítio catalítico e a instabilidade de alguns grupos químicos da cadeia em formação irão influir em ciclizações espontâneas na cadeia (WEISSMAN, 2009; BRACHMANN *et al.*, 2007).

1.3.11. Origem das PKSs

No final de 1980, ficou evidente que a despeito da sua vasta heterogeneidade estrutural, todos os policetídeos são obtidos por um processo intimamente semelhante aos da formação dos ácidos graxos pelas FASs (“Fatty Acid Synthases”), havendo evidências do compartilhamento de enzimas homólogas entre as FASs e as PKSs. Derivados simples de ácido carboxílico (como acetil-CoA, malonil-CoA e metil-malonil-CoA) são concatenados em cadeias lineares por uma condensação tipo “Claisen”, seguido em muitos casos por modificações redutivas em seus grupos β -cetônicos (WEISSMAN, 2009).

No entanto, a base química da programação de modos alternativos destas sintases para as transformações sobre o arcabouço químico policetônico geral produzido pela ligação tipo “Claisen” e a construções das moléculas policetonas (não reduzidas); dos policetídeos (parcialmente reduzidos); ou dos “Poly-unsaturated Fatty Acids” (PUFAs-totalmente reduzidos), permanece obscura. Resultados de pesquisas envolvendo estudos bioquímicos, sequenciamento de centenas de genes codificantes para PKSs e enzimas de modificação pós síntese, tem oferecido o caminho mais promissor para a compreensão destas vias bioquímicas, esclarecendo os passos pelos quais são gerados cada grupo químico em algumas vias (Fig. 12). Contudo, ainda não está disponível nenhum mecanismo óbvio para explicar as escolhas do tipo e do número de monômeros empregados, bem como a ordem em que os mesmos são incorporados, e a extensão do potencial redox ajustado a cada etapa da extensão da cadeia, além do perfil de ciclização e lactonização da cadeia resultante (WEISSMAN, 2009).

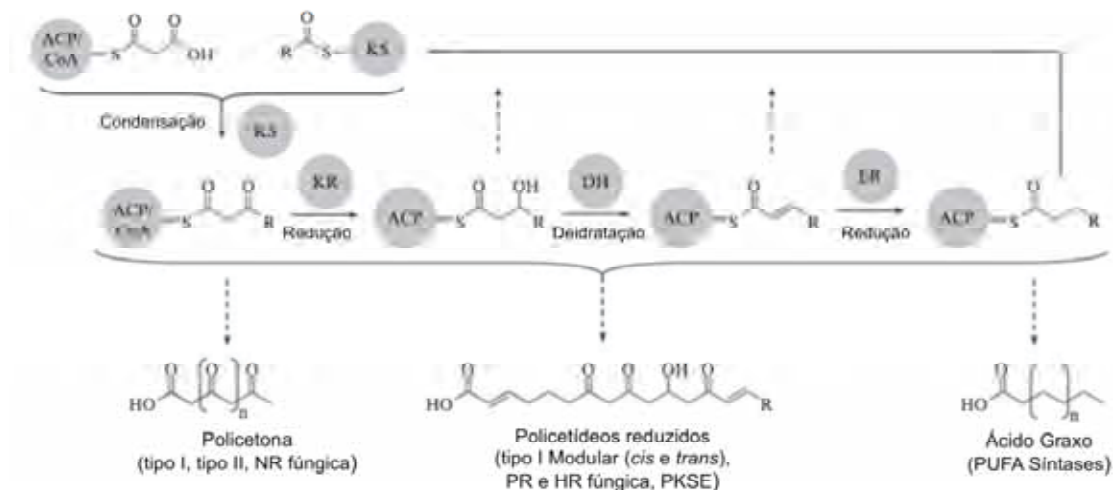


Figura 12. Esquema genérico das reações para biossíntese tanto de ácido graxo quanto de policetídeos, ilustrando como o ciclo básico de reação da biossíntese de ácidos graxos pode ser desviada (“shunted”) para gerar diversas estruturas policetílicas, incorporando diferentes níveis de redução, e guardando-se as particularidades de cada via (WEISSMAN, 2009).

Estudos envolvendo filogenia (METSÄ-KETELÄ et al., 1999) e cristalografia de raio X (CASTALDO et al. 2008) têm lançado luz a este impasse, destacando a contribuição da KSA (umas das enzimas do cluster de PKS) como uma das responsáveis para a escolha do substrato empregado na síntese. Contudo, os mecanismos que levam ao controle do tamanho da cadeia química parecem envolver tanto a natureza das enzimas da *pks mínima* (SHEN, 1999), quanto a estereoespecificidade de outras enzimas, tais como as aroclases-didomínio (ZAWADA E KHOSLA, 1997).

Essas similaridades no conteúdo gênico, bem como semelhanças gerais na reações enzimáticas entre as FAS e PKS, deriva de um parentesco filogenético, como pode ser observado na Figura 13 (JENKE-KODAMA e DITTMANN, 2009). Os autores compararam tanto FAS quanto FABs (animais, fungos e bactérias e de organelas celulares, além de PKS I envolvidas com a síntese de glicolipídeos e PUFAs) às PKS II de bactérias, e aos vários tipos de PKS I de bactérias e fungos. Os autores utilizaram as FabH de arqueobactéria e de bactérias como grupos externos no estudo, por serem

considerados os homólogos mais distantes do grupo. Na hipótese mais verossímil, as KSA de PKS II (bacteriana) surgiram em paralelo com as proteínas FabF de FASs bacterianas (ramo 3, Fig. 13). As FASs do tipo bacterianas também são encontradas em mitocôndrias de eucariotos superiores. Dois grupos de FASs parecem representar a mais antiga fusão entre formas de enzimas bacterianas de FAS e PKS originalmente distintas (ramo 5, Fig. 13). As FASs de fungos (ramo 5, Fig. 13) são parafiléticas e distantes das FASs de animais (ramo 10, Fig. 13), e revelam diferenças nas ordens de seus domínios enzimáticos e na estrutura tridimensional das proteínas.

A topologia da árvore permite visualizar uma hipótese de homologia para o surgimento das PKSs. É possível verificar que as policetídeos sintases são um grupo parafilético, sendo que cada família distinta compartilha um ancestral comum com uma determinada família de ácido graxo sintases ao invés de se relacionar mais diretamente com as outras KSA de PKSs (JENKE-KODAMA e DITTMANN, 2009). Trata-se também de um grupo polifilético, surgindo várias vezes ao longo da evolução, sempre em proximidade com algum grupo de FAS, o que explica a grande proximidade das ácido graxo sintases, ao mesmo tempo em que apresentam tantas diferenças de uma família para outra.

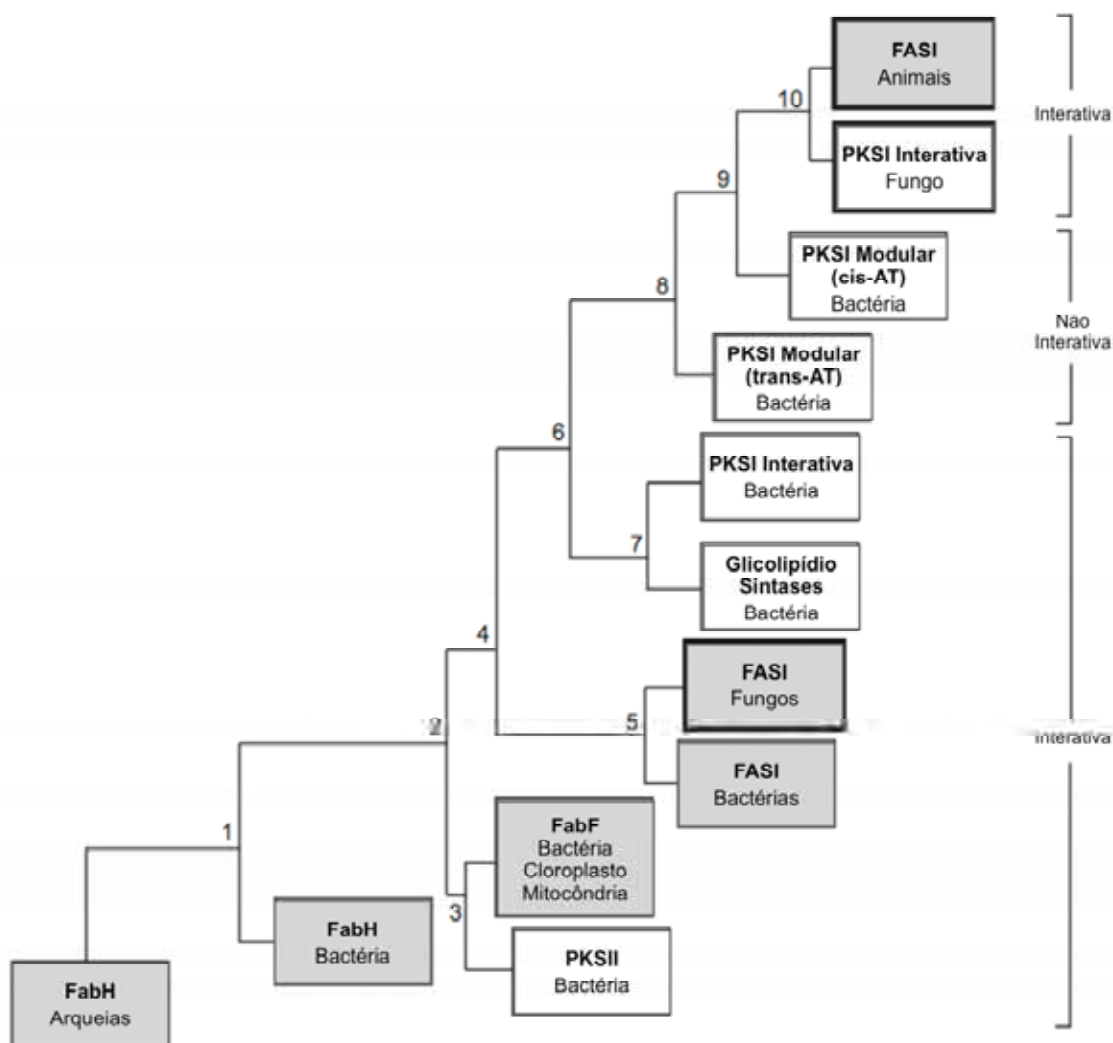


Figura 13. Comparação filogenética entre FAS e PKSs (JENKE-KODAMA e DITTMANN, 2009).

Tendo em vista a importância dos ácidos graxos no metabolismo, indo desde molécula estrutural de membrana, para reserva energética e precursores de sinais bioquímicos, era de se esperar que o grupo enzimático tivesse um caráter mais conservado. Contudo, depara-se com uma grande diversidade de ácido graxo sintases, ao mesmo tempo em que existe uma semelhança tão grande entre estas e as PKSs que indicam muitos pontos de intercâmbio gênico entre os diversos grupos enzimáticos das duas rotas. A hipótese mais plausível é a da ocorrência de Transferência Horizontal

Gênica (HGT), sendo encontradas evidências diretas do fenômeno entre fungos e bactérias. Estudos revelaram também que as PKS III, tão difundidas entre os vegetais são na verdade originadas de transferências horizontais de bactérias para plantas, ao contrário da hipótese original, em que se acreditava que as plantas seriam as primeiras detentoras das vias de chalconas sintases (GROSS et al., 2006; JENKE-KODAMA e DITTMANN, 2009).

Uma análise feita sobre 17 “clusters” de PKS I previamente caracterizados para *Streptomyces* revelou que a maioria dos agrupamentos surgiram por meio de eventos de duplicação gênica (JENKE-KODAMA e DITTMANN, 2009). Outro evento que causa diversificação é a recombinação intra domínios (PIDOT et al., 2008), possibilitando a um mesmo organismo a produção de vários compostos através de recombinação de módulos de uma determinada via, como substituição de Acil-transferases (AT) em módulos de ativação (“Loading Module”), bem como a perda de módulos de extensão. Genes de vias distintas também podem se recombinar entre si, havendo relatos de substituições de módulos inteiros (JENKE-KODAMA e DITTMANN, 2009). Em PKS I, a substituição de um domínio AT pode levar à diferenciação da própria especificidade ao substrato na via, o que indica que novos produtos podem surgir uma vez que mudam-se os precursores. Da mesma forma, todos esses eventos podem ocorrer para as vias de PKS II e vias híbridas de PKS/NRPS (“Non-ribosomal peptide Synthetase”).

Como se não bastasse as possibilidades de recombinação intragenômicas, os eventos comuns de HGT em *Streptomyces* (METSÄ-KETELÄ et al., 1999) magnificam esse efeito recombinatório em muitas ordens de grandeza, justificando a imensa variação dos compostos policetídeos ao longo de milhares de anos de evolução.

FISCHBACH, WALSH & CLARDY (2008) descrevem ainda como se dá outro evento de diversificação e propagação das PKSs entre vários grupos de micro-organismos: a **evolução coletiva** de genes. Os autores indicam que os “clusters” microbianos são os elementos gênicos que mais rapidamente sofrem evolução, sujeitos tanto a eventos comuns de transferências horizontais quanto a transferência vertical mais frequente dado o curto tempo de geração destes organismos. Em virtude do

rápido diagnóstico de mudança fenotípica por mudanças de vias, o grupo também assume importância como modelo de estudo da evolução conjunta de genes.

Um fator decisivo para a propagação das vias de policetídeos é o fato de que muitas vezes estas pequenas moléculas irão conferir vantagens adaptativas ao seu hospedeiro, e muitas vezes são propagados por plasmídeos. O mecanismo que mantém a evolução dos policetídeos é o elevado custo de manutenção destas vias pelos organismos que as possuem: a partir do momento que a mesma não confere vantagem adaptativa para o organismo hospedeiro, torna o seu tempo de vida evolutivo bastante curto (FISCHBACH, WALSH & CLARDY 2008). Essa seleção negativa (aumento do gasto metabólico), aliada à seleção positiva por vantagens adaptativas conferidas por vias bem sucedidas, criam um crivo muito eficaz para a inovação de moléculas bioativas, ampliando muito o leque de possibilidades para o campo das pesquisas de PKS.

1.3.12. As PKS II

Um grupo de três enzimas, comumente chamado de Policetídeo Sintase Mínima ou PKS mínima, é compartilhado por todas as vias da Família PKS do tipo II: a Cetoacil-Sintase (KS α), um Fator Elongador de Cadeia (KS β) e a ACP, uma Proteína Carreadora de Acila (WAWRIK et al., 2005). As moléculas policetílicas produzidas por PKSs do tipo II são normalmente glicosiladas, e as alterações no padrão destas glicosilações geram uma diversidade adicional a este grupo (HOPWOOD, 1997). Outros genes conferem a diversidade na estrutura da cadeia policetílica e podem estar presentes em alguns organismos e ausentes em outros (HUTCHINSON & FUJJI, 1995).

Os policetídeos aromáticos constituem uma importante família de bioativos de emprego farmacológico, muitos dos quais com ação antibiótica e anti-câncer, e incluem as tetraciclinas, antraciclinas, ácidos aureólicos, tetracenomicinas, anguciclinas, benzo-iso-cromoquinonas, e os policetídeos mais conhecidos são os produzidos pelo metabolismo dos *Actinomycetes* (ZHANG e TANG, 2009).

A despeito da prospecção por novos *clusters* de PKS II objetivar frequentemente novos antibióticos, essa é apenas uma descrição parcial ou imprecisa de sua função, uma vez que podem estar envolvidos na síntese de pigmentos de esporo, na diferenciação fisiológica em condições ambientais adversas (JENKE-KODAMA e DITTMANN, 2009) e mesmo na degradação de compostos lignínico-celulósicos (HSIAO e KIRBY, 2008).

Por outro lado, mesmo trabalhos empregando expressão de genes relacionados a produção de pigmentos demonstram o quanto a biossíntese combinatória destes “clusters” pode contribuir com a farmacologia. Como exemplo, SHEN (1999) descreveu um interessante trabalho em que organismos recombinantes foram produzidos a partir da introdução de uma *pks mínima* derivada de PKS II de esporo de pigmento, resultando em uma variedade de novos esqueletos carbônicos muito superior a encontrada em ensaios que empregaram a *pks mínima* de produtores naturais de antibióticos. Um composto policetídeo componente de pigmento pode ainda apresentar atividade farmacológica “*in natura*”. Alguns podem ser antitumorais ou antibióticos fracos, como é o caso de alguns pigmentos antraquinônicos (BRACHMANN et al. 2007). Os pigmentos do tipo nafitoquinônicos (policetídeos de origem fúngica) são conhecidos por sua atividade antibacteriana, antifúngica e inseticida (GRAZIANI et al. 2004).

Segundo HSIAO e KIRBY (2008), o gênero *Streptomyces* ocupa um importante papel no solo, atuando na reciclagem dos materiais de natureza lignínico-celulósica, e algumas enzimas envolvidas no processo parecem derivar de vias de policetídeos de pigmento. A determinação de quais e quantas enzimas, precisamente, estariam envolvidas na rota metabólica de degradação de celulose e lignína (incluindo genes não relacionados a PKS) permanece obscura, mas é possível evidenciar a participação de: catalases/peroxidases; endoglucanases; celulasas; proteínas ligantes de celulase; mono-oxigenases; dioxigenases; hemeoxigenases; tirosinases (monofenol/mono-oxigenases) e seus cofatores; celobiose-hidrolases; oxigenase; alcanos-mono-oxigenases; e reguladores sensores de peróxido de hidrogênio, além de enzimas de metabolismo de celobiose. Algumas ainda podem também se relacionar novamente às

vias de PKS de forma indireta, através do fornecimento de unidades precursoras para modificação da cadeia de policetídeos.

1.3.13. Biossíntese combinatória

A formação em “*cluster*” dos genes para as PKS II favorece a prospecção por PCR de novas rotas biossintéticas em bibliotecas metagenômicas. Isso se justifica pelo fato de que localizar a sequência conservada de uma única região gênica em um dado clone metagenômico pode implicar em encontrar não apenas a enzima para qual o oligonucleotídeo iniciador foi desenhado, mas todas as outras que estão localizadas nas adjacências suportadas pelo tamanho do inserto do clone. Dessa forma, existem grandes chances de recuperação de uma via biossintética na íntegra, incluindo promotores, reguladores, genes que conferem resistência ao hospedeiro heterólogo para o antibiótico em questão, enzimas de modificação pós-síntese da estrutura policetídica, dentre outros.

Além da abordagem clássica de prospecção de fármacos que se fundamenta na descoberta de moléculas naturais, outra estratégia na busca por novos fármacos seria o uso da engenharia molecular. Essa engloba o conjunto de manipulações gênicas inerentes às práticas de expressão heteróloga e de biossíntese combinatória, embora não se restrinja à apenas essas abordagens.

No caso da expressão gênica heteróloga, um exemplo de emprego da estratégia seria a introdução de uma maquinaria gênica responsável por determinada molécula em um hospedeiro mais eficiente do que o original para a produção do composto em escala industrial (MARTIN & DEMAINE, 1980). Outra variação do emprego dessa técnica ganha destaque a partir da expansão do uso de abordagens metagenômicas, nas quais muitas vezes o material gênico promissor encontrado por homologia pertence a um organismo desconhecido, sem possibilidades de seu isolamento e cultivo laboratorial pelas técnicas disponíveis. Nesse caso, a expressão heteróloga se configura como a via mais

acessível para tentativa de expressão e obtenção dos compostos enzimáticos ou químicos de interesse.

Um fator importante para as atividades de prospecção por novas moléculas, utilizando expressão heteróloga, é o fato de atualmente o pesquisador dispor do acesso às previsões parciais da estrutura tridimensional de enzimas, e mesmo de estruturas químicas de compostos que possam ser produzidos por maquinarias bioquímicas similares (BALTZ, 2006). Essa previsão pode ser obtida por representações *in silico* criadas com o auxílio de ferramentas de bioinformática e da engenharia molecular, a partir da sequência nucleotídica estudada, e das informações reunidas pelo pesquisador em relação à família enzimática de interesse.

A partir do “esboço” molecular do candidato a fármaco são fornecidas as pistas sobre a relevância deste metabólito, através de uma comparação dos prováveis grupos químicos com estruturas similares conhecidas. Indo um pouco além, algumas vezes é possível uma formulação de hipóteses sobre a interação de tal composto com células vivas (bioatividade) por meio da comparação de suas características com bancos de moléculas com funções já elucidadas. Uma vez que se tem uma idéia de quão promissor um determinado “cluster” seja, a etapa seguinte seria a introdução deste em hospedeiros diferentes para testes de expressão. No caso específico de genes para o gênero *Streptomyces*, existe a disponibilidade para a pesquisa de hospedeiros heterólogos aptos a receber o DNA exógeno, com vantagens de serem selecionados e modificados para não silenciar a expressão de genes heterólogos. Como exemplos, têm-se linhagens de *Streptomyces lividans*, *Streptomyces albus* e *Streptomyces coelicolor*, além de linhagens mutantes de *Escherichia coli* desenvolvidas especialmente para expressão de genes do grupo de *Streptomyces* (MARTIN & DEMAINE, 1980; WANG et al., 2000; BALTZ, 2006).

Por sua vez, a biossíntese combinatória voltada para a busca de novos fármacos também é um campo de abordagem para pesquisadores que trabalham em Metagenômica.

A produção de um dado composto policetídeo específico é resultado da interação enzimática peculiar a um grupo de organismos que compartilhem de uma

mesma rota bioquímica para tal metabólito secundário. Para certos tipos de antibióticos, existem organismos produtores do mesmo composto entre linhagens pertencentes a espécies distintas. Existem duas situações que exemplificam o compartilhamento do mesmo metabólito por um dado conjunto de organismos. Um dos exemplos é o que ocorre dentre espécies filogeneticamente próximas entre si, que porventura possuam uma mesma via biossintética herdada de um grupo ancestral comum. A outra possibilidade se sucede dentre indivíduos de relativo distanciamento filogenético, mas que, no entanto, tenham incorporado “clusters” enzimáticos do outro organismo por transferência gênica horizontal. Esta troca de genes entre organismos não detentores de relações íntimas de parentesco se constituiu em um fenômeno de relatada ocorrência dentro do gênero *Streptomyces* (METSÄ-KETELÄ *et al.*, 1999).

No caso de antibióticos da PKSs, o emprego da biossintese combinatória se beneficia do fato de que, embora cada rota bioquímica normalmente resulte em um tipo específico de moléculas, irão existir dentro de uma mesma família de PKS enzimas comuns a todas as rotas, derivadas de genes homólogos e detentores de regiões conservadas, ainda que não sejam idênticas entre si.

Em virtude desta semelhança enzimática existirão etapas bioquímicas similares nos mecanismos básicos envolvidos na formação da cadeia policetílica. Esta conservação de certas propriedades dentre as enzimas homólogas permite certa compatibilidade entre grupos enzimáticos de “clusters” diferentes. No caso de antibióticos como os da Família PKS II, em que a produção da molécula final é resultado da interação de várias enzimas pertencentes a um “cluster”, a biossintese combinatória permite criar moléculas, através da recombinação de módulos enzimáticos de organismos produtores já conhecidos, bem como combinações de módulos conhecidos com genes novos obtidos pela prospecção em metagenômica.

Assim, a engenharia molecular ou biossintese combinatória engloba um conjunto de manipulações da genética molecular e outras manipulações biossintéticas, que basicamente podem ser entendidos como estratégias para acelerar os processos evolutivos em muitas ordens de magnitude visando “competir” com a evolução natural dos antibióticos. Como exemplo deve-se considerar a imensa quantidade de complexos

peptídeos, policetídeos e outros arcabouços químicos conhecidos, amplamente diversificados pela presença e disposição de estruturas tais como: glicosilações, metilações, halogenizações, grupos acilas e hidroxilações, e que se desenvolveram ao longo de bilhões de anos em actinomicetos e fungos (BALTZ, 2006).

1.3.14. Breve revisão de Estruturas de Proteínas

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. Se se trata de uma enzima, esse correto empacotamento é importante para a construção do ambiente estereoquímico do sítio catalítico envolvido na atuação sobre as inúmeras reações bioquímicas. Se a proteína é estrutural, esse dobramento lhe conferirá a forma correta necessária para sua função: fibrosas, globulares, e as características químicas que permitirão a correta localização, como no caso de proteínas constituintes de transportadores de membrana celular, que graças a diferenciação entre porções hidrofílicas e hidrofóbicas podem atuar como unidades protéicas transmembrânicas.

A estrutura primária (Fig 14) 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 se dá graças 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 irão 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”. Essa atração é responsável pela estruturação de regiões proximamente localizadas na cadeia em

elementos de forma específicas, denominadas de estruturas secundárias, tais como as α - ou β - hélices, folhas- β , “coils” e “loops”.

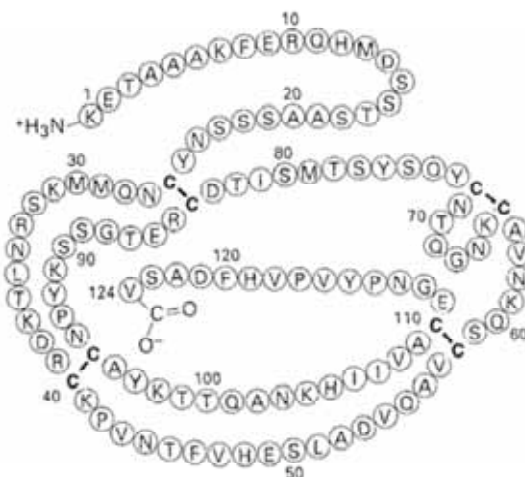


Figura 14. Esquema da estrutura primária da ribonuclease bovina, ilustrando o início das interações que darão origem a estrutura secundária. (Fonte: WWW.SWISS MODEL)

A Figura 15.A ilustra a presença de 3 estruturas secundárias do tipo α - hélices, ligadas entre si por “loops”. Estas estruturas secundárias por sua vez, também interagem entre si graças a interações eletrostáticas, hidrofóbicas, hidrofílicas, pontes de dissulfeto, e atuação de forças de “Wan der waals”, criando formas mais complexas de dobramento, que aproximam regiões distantes da cadeia, criando a estrutura terciária protéica (Fig. 15.B)

A interação entre duas ou mais cadeias polipeptídicas estruturadas na forma de dobramento terciário dá origem a estrutura protéica quaternária, como pode ser observado na (Fig. 15.C)

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.

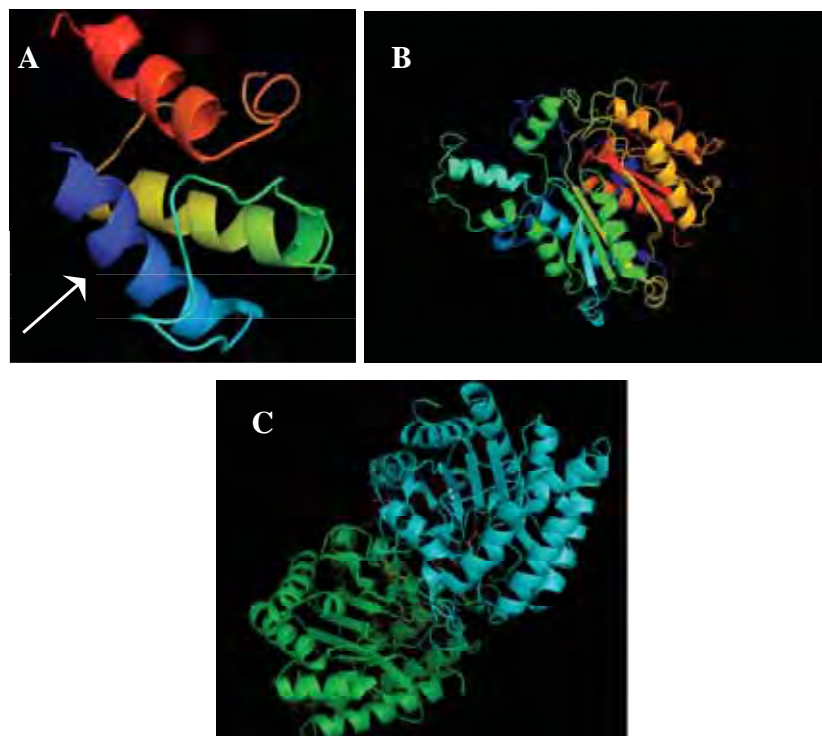


Figura 15. Esquema representativo de estruturas protéicas. **A:** Evidência de Estruturas 2D, representadas na figura por alfa-hélices (estrutura 2D apontada pela seta) e pelos “coils” de união. **B:** Exemplo de estrutura 3D, onde é possível evidenciar a organização de motivos em alfa-hélices, folhas-beta, “coils” e “loops”. **C:** Estrutura quaternária, com representação de duas cadeias: uma em azul outra em verde, para constituição da molécula funcional.

1.3.15. Modelagem enzimática *in silico* e avaliação (“assessing”) de estruturas 3D

A despeito do grande número de sequências protéicas depositadas em bancos públicos de dados, o qual ultrapassava seis milhões, até o ano de 2009 estima-se que apenas 50 mil proteínas tiveram sua estrutura experimentalmente determinada. Isso reflete o quão difícil, demoradas e dispendiosas são tais técnicas, a despeito de sua importância (KELLEY & STERNBERG, 2009). Um método alternativo a essas técnicas, que vêm auxiliar os pesquisadores na predição de substratos e reconhecimento de sítios ativos, é a modelagem 3D *in silico* de proteínas. Essas abordagens exploram o princípio de que embora as enzimas e proteínas naturais sejam extensivamente diversas, os padrões de arquitetura e dobramento não são, de forma que as mais

distantemente homólogas proteínas tendem a desenvolver um tipo similar de estrutura (KELLEY & STERNBERG, 2009). Avanços no desenvolvimento de programas de modelagem de estrutural, tais como “PSI-Blast”, “Hidden Markov Model” (Modelo Oculto de Markov) e “Profile–profile matching algorithms” (Algoritmos de sobreposição perfil-á-perfil) permitem uma predição e modelagem relativamente acurada de novas proteínas com menos de 20% de identidade com uma proteína de estrutura conhecida, demonstrando o quão poderosas e promissoras estas abordagens são para o estudo de novas rotas biossintéticas (KELLEY & STERNBERG, 2009; BALTZ, 2006). Ao contrário dos primeiros modelos, os quais eram limitados à predição de estruturas pequenas (menos de 100 resíduos de aminoácidos) com base nas características químicas de sua estrutura primária, estes realizam a modelagem com base em moldes homólogos e reconhecimento de formas (“fold-recognition”). As aplicações da predição da estrutura terciária das proteínas abrangem a formação de hipótese funcional para proteínas hipotéticas, proporcionam sinais de fase em cristalografia, permitem a seleção de sítios para a mutagênese e o desenho racional de novas drogas (KELLEY & STERNBERG, 2009). Embora as predições para sequências com similaridade abaixo de 20% sejam possíveis, a confiabilidade da modelagem é diretamente proporcional a identidade residual da proteína alvo e do molde (“template”). Valores acima de 40% de identidade retornam modelos preditivos de alta confiabilidade, ideais para estratégias de predição de ligantes em sítios ativos, como drogas ou mesmo substratos (“docking”). Valores abaixo, mas acima de 30%, são empregues para estratégias tais como inferências evolutivas. Abaixo de 30%, os modelos apresentam muitas falhas e extensões de gaps e erros de alinhamento, e passam a ser modelos preditivos limitados, mas de utilização para verificação geral de padrões aproximados de estruturas desconhecidas. Para melhorar a possibilidade de predição em estruturas para as quais existam um valor limitado de moldes, o cuidado com a escolha destes, a precisão no alinhamento global e o emprego de alinhamentos múltiplos ajudam a melhorar a acurácia de predição dos modelos (DA SILVEIRA, 2005). O terceiro método explora a modelagem de sequências por homologia, em que primeiro se faz o alinhamento entre sequências.

Após confecção dos modelos estes devem ser checados para verificar o padrão de confiabilidade em relação às estruturas reais. Para tanto, é importante checar a conservação dos resíduos importantes para o sítio ativo protéico, verificando a correspondência dos padrões destas regiões entre o modelo e o molde. Outros dados muito importantes dizem respeito à boa qualidade estereoquímica do modelo, refletida pelos padrões adequados de comprimento de ligação, ligação peptídica e planaridade 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. Tais características podem ser obtidas pelo uso dos programas PROCHECK (LASKOWSKI et al., 1993). Além disso, estes programas possibilitam a visualização do perfil de qualidade estereoquímica com base no gráfico de Ramachandran, que permite visualizar a disposição espacial do modelo em regiões mais favoráveis, com base na conformação molecular com menor energia livre no sistema. 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, tais como o VERIFY3D, PROSAIL e o HARMONY (DA SILVEIRA, 2005). Valores de R.M.S.D. (programa CCP4 e VERIFY3D) auxiliam a identificar a 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. Tomados em conjunto, estes valores irão indicar quais as informações biológicas que podem ser inferidas a partir de um dado modelo, garantido que as extrapolações feitas a partir dos resultados assumam proporções confiáveis, limitados à capacidade preditiva do modelo.

CAPÍTULO 2 – Análise funcional por homologia de um “cluster” de PKSII metagenômico

Elisângela Soares Gomes, Viviane Schuch, Gabriel Padilla, Luciano Takeshi Kishi, Silvana Giuliatti, Eliana Gertrudes de Macedo Lemos

RESUMO - Policetídeos aromáticos são metabólitos secundários diversos presentes em algumas espécies de micro-organismos e plantas. Constituem um vasto número de moléculas de reconhecida importância ecológica e farmacológica, sintetizadas por complexos enzimáticos denominados PKSII. Excetuando-se o *core* biossintético ***pks mínima***, estes agrupamentos assumem combinações enzimáticas imprevisíveis relacionadas à modificação das cadeias policetílicas durante e depois de sua síntese. O objetivo deste trabalho foi o acesso à novas vias de policetídeos empregando oligonucleotídeos degenerados para prospecção em uma biblioteca metagenômica. Um inserto de 22.988 pares de bases foi encontrado e sequenciado. A predição funcional das ORFs metagenômicas por homologia e a identificação de domínios enzimáticos resultaram em candidatos a genes relacionados à produção de policetídeos e, possivelmente, envolvidos com degradação de compostos lignínico-celulósicos. Dentre estes, foram encontrados homólogos à: transportadores ABC; *PKS mínima*; Ciclases; Aromatases; Hidroxilases; Mono-oxigenases; O-metil-transferases; Lacases; β -glicosidases; Celulases; Peptidases; Acil-transferases; além de reguladores gênicos Lacl e LuxR. Foi feita a caracterização filogenética do “cluster” metagenômica através da filogenia da enzima diretamente envolvida na escolha de substratos (KSA), bem como a comparação entre as enzimas do “cluster” com a arquitetura de outras vias de PKSII. O “cluster” possui uma arquitetura similar a agrupamentos envolvidos na síntese de pigmentos, porém apresenta uma maior complexidade.

Palavras-chaves: PKSII, KSA, Metagenoma, Degradação de lignina/celulose.

2.2. Introdução

Além de atuarem em todos os processos de ciclagem, manutenção, expansão da vida no planeta, os micro-organismos representam em antítese os principais agentes etiológicos das doenças conhecidas e a principal fonte de compostos bioativos explorados pela farmacologia. Dentre os compostos naturais conhecidos, os policetídeos são responsáveis por um faturamento anual superior a U\$ 20 bilhões, e mais de um terço do total das novas drogas implementadas no período de 2005 a 2007, ocupando um lugar de destaque (WEISSMAN, 2009).

Embora os números sejam expressivos, nos últimos 30 anos apenas duas classes de antibióticos inéditas foram lançadas no mercado: a oxazolidinona linezolida e o lipopeptídeo cíclico daptomicina. Como agravante, segundo a “*US Food and Drug Administration*” (FDA), a telavancina, a retapamulina e a tigeciclina foram as únicas moléculas totalmente novas liberadas nos últimos cinco anos, contrastando com a rapidez em que os antibióticos em circulação se tornam ineficientes (HAMAD, 2010).

Atendendo a crescente demanda da farmacologia frente à plasticidade de adaptação dos patógenos, os policetídeos têm sido alvos da engenharia molecular desde 1985 para desenho de compostos bioativos “sob medida”: mais eficientes, menos tóxicos ou inéditos, através da biossíntese combinatória (HOPWOOD *et al.*, 1985; RIX *et al.*, 2002).

As perspectivas atuais dessas abordagens são ainda melhores, uma vez que se vive a era das ciências “ômicas” (metagenômica, proteômica e metabolômica), quando não mais se cerceia o alcance de biomoléculas pela habilidade de cultivo laboratorial do micro-organismo produtor, a qual seria de apenas 1% (AMMAN, LUDWING e SCHLEIFER, 1995; SCHMEISSER, *et al.* 2007; COMMITTEE ON METAGENOMICS, 2007).

O complexo de enzimas que estão relacionadas à produção de policetídeos é dividido em três grandes famílias homólogas às Ácido Graxo Síntase (FAS): as PKSs dos tipos I (PKS Modular/Interativas), II (PKS aromática) e III (Chalcona Síntases),

podendo também haver subfamílias e “clusters” híbridos (SHEN, 2003; WEISSMAN, 2009; JENKE-KODAMA & DITTMANN, 2009).

Os policetídeos são, originalmente, produtos do metabolismo secundário de algumas espécies de bactérias, fungos, plantas e mesmo microalgas como dinoflagelados, e compõem a estrutura química de muitos fármacos (RIX *et al.*, 2002). Dentre estes, os diferentes padrões de glicosilações e hidroxilações, e a facilidade em obtenção de vias recombinadas fazem da PKS II um alvo bastante promissor para a farmacologia.

A despeito da prospecção por novos “clusters” de PKS II objetivar, frequentemente, novos antibióticos, essa é apenas uma descrição parcial ou imprecisa de sua função, uma vez que podem estar envolvidos na síntese de pigmentos de esporo, na diferenciação fisiológica em condições ambientais adversas (JENKE-KODAMA & DITTMANN, 2009) e mesmo na degradação de compostos lignínico-celulósicos (HSIAO & KIRBY, 2008). Por outro lado, um pigmento policetídeo pode apresentar atividade farmacológica “*in natura*” (BRACHMANN *et al.* 2007; GRAZIANI *et al.*, 2004). Além disso, a expressão de genes de “clusters” de pigmentos demonstra o quanto a biossíntese combinatória pode contribuir com novas moléculas policetílicas na busca por novos fármacos (SHEN, 1999).

Neste trabalho descreve-se os esforços empreendidos no sentido de se obter genes relacionados à família PKSs aromáticas (PKS II), empregando uma biblioteca metagenômica descrita previamente (SCHUCH *et al.* 2009). Para estudo de uma sequência encontrada foram empregues análises por homologia de ORFs e Domínios enzimáticos, análise filogenética de cetosintase alfa (KSA) e comparação da arquitetura de “clusters”.

2.3. Material e Métodos

Biblioteca

A biblioteca cosmidial utilizada foi previamente construída (SCHUCH, *et al.* 2009) utilizando o kit *pWEB-TNC Cosmid Cloning* (Epicentre Biotechnologies Madison, WI,

USA) de acordo com as instruções do fabricante. A biblioteca é constituída de clones comportando grandes insertos (20-45 kb), obtidos pela clonagem de fragmentos de DNA ambiental extraído de um solo sob serrapilheira de bosques de eucaliptos não sujeito a nenhum trato cultural, ou revolvimento de substrato por atividade humana por cerca de 40 anos (SILVEIRA et al. 2006).

Prospecção de agrupamentos gênicos de PKSII

A prospecção de genes para PKS II foi conduzida por meio de amplificação por PCR utilizando oligonucleotídeos degenerados complementares a uma região altamente conservada do “cluster”: a KSA (ou KS α) previamente desenhados (METSÄ-KETELÄ et al., 1999). Para a padronização da reação foram empregues amostras de DNA gênomico de *Streptomyces chartreuses*, *S. rimosus* e *S. galilaeus*, extraídas pelo método de Marmur (1961).

As melhores condições de amplificação foram obtidas para as seguintes concentrações de reagentes: 50-100 ng de DNA genômico; 40 pmoles de cada oligonucleotídeo; 1 mM de dNTPs; 2,5% de DMSO; 2 μ l de tampão de reação 10x concentrado; MgCl₂ 2,25 mM; 2U de Taq DNA polimerase (Invitrogen) para um volume final de reação de 20 μ l. A reação foi conduzida em um termociclador automático, com as seguintes temperaturas: 2 minutos de desnaturação inicial a 96°C; 30 ciclos de: 1 minuto de desnaturação a 96°C, 1 minuto de hibridização com a temperatura 64°C, 1,5 minutos de alongação a 73°C; e uma extensão final de 8,5 minutos à 73°C.

Para a prospecção, as amostras dos clones foram reunidas em “pools” de 96 clones antes da extração de DNA dos cosmídeos pelo método de lise alcalina (SAMBROOK e RUSSEL, 2001), totalizando 94 “pools” de amostras constituídas por 96 clones cada. A amostras compostas de DNA foram então empregues como molde para as reações de amplificação por PCR. Somente os “pools” que apresentaram amplificação positiva foram submetidos a uma extração individual de DNA cosmidial, seguido de amplificação por PCR para identificar quais os clones com amplificação

positiva. Os “amplicons” foram purificados usando o *kit GFX PCR DNA and Gel Band Purification* (Amersham Biosciences, São Paulo, Brasil), seguido de clonagem em vetor *pGem T- Easy* (Promega), e transformação dos plasmídeos recombinantes para o hospedeiro heterólogo *Escherichia coli DH5 α F⁻* (supE44, Δ lacU169 (ϕ 80 lacZ Δ M15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1) sensível a ampicilina. Os clones obtidos foram submetidos a extração pelo método de lise alcalina (SAMBROOK e RUSSEL, 2001). O inserto dos clones foi sequenciado com os oligonucleotídeos universais do vetor (Sp6 e T7), usando *DYEnamic ET Terminator* (Amersham Biosciences, São Paulo, Brasil). A eletroforese foi conduzida em sequenciador capilar automático *ABI 3700* (Applied Biosystems, São Paulo, Brasil) de acordo com as instruções do fabricante. As sequências no formato FASTA foram submetidas ao GenBank do “National Center for Biotechnology Information” (NCBI) para a comparação à sequências homólogas depositadas utilizando a ferramenta BLASTn, e escolha do cosmídeo comportando uma KSA. Uma vez escolhido o clone cosmidial, o tamanho de seu inserto foi estimado através de restrição enzimática, seguida por uma separação dos fragmentos por meio de eletroforese em gel de agarose (1,5%) e estimativa do tamanho molecular de cada fragmento de restrição tendo por base um padrão de tamanho molecular comercial (“1Kb DNA ladder”, *Fermentas*). O tamanho do inserto foi calculado subtraindo-se o tamanho molecular conhecido para o vetor não clonado do valor total obtido para a soma dos fragmentos de restrição.

Subclonagem, sequenciamento do cosmídeo escolhido e análises de bioinformática

O sequenciamento do clone cosmidial selecionado foi conduzido utilizando-se a técnica de biblioteca “*shotgun*”. O clone foi submetido a nebulização por 12 segundos à pressão com nitrogênio gasoso de 3 Kg/cm². o DNA nebulizado foi submetido a uma eletroforese preparativa, na qual foram selecionados fragmentos entre 1,5 à 2 kb de extensão. Estes foram clonados em vetores *pUC19DNA/SmaI* (Tools), obtendo-se 10 placas de 96 clones cada. O DNA plasmidial foi obtido através do método de extração

por lise alcalina (SAMBROOK e RUSSEL, 2001). Os clones da sub-biblioteca foram submetidos à PCR de sequenciamento utilizando o *DYEnamic ET Terminator* (Amersham Biosciences, São Paulo, Brasil) com os oligonucleotídeos M13F e M13R.

A eletroforese foi conduzida conforme descrito no ítem anterior. Os dados brutos dos eletroferogramas foram analisados utilizando o pacote de programas Phred/Phrap (EWING *et al.*, 1988; EWING e GREEN, 1998) e Consed (GORDON *et al.*, 1998) para gerar as sequências FASTA e avaliar a qualidade das sequências individuais (“reads”) e das resultantes da montagem (“assembly”). Valores de Phred acima de 20 foram considerados de boa qualidade. A sequência final para o inserto (“consensus”) foi obtida através do desenho de oligonucleotídeos de extensão (“*primers walking*”) para unir os dois maiores fragmentos de montagem (“contigs”) gerados pelo algoritmo de montagem (“*asssembler*”) para que se pudesse superar uma barreira de sequenciamento contida no inserto cosmidial ocasionada por um alto conteúdo C/G.

Foram identificadas as Fases Abertas de Leitura (ORFs - *software* ORFfinder: <http://www.ncbi.nlm.nih.gov/projects/gorf/>) no inserto metagenômico, e foi feita uma triagem manual de prováveis genes com base na maior similaridade a sequências depositadas em bancos públicos de dados integrados ao NCBI (BlastP, ALTSCHUL *et al.*, 1990), seguido de correção de erros da anotação em códons de iniciação. A mesma sequência nucleotídica do “*contig*” foi submetida a um programa de identificação de domínios conservados: o PRODOM (SERVANT *et al.*, 2002), buscando uma confirmação da previsão funcional. As ORFs foram também comparadas às entradas do PFAM (Banco de famílias enzimáticas: FINN *et al.*, 2010) e submetidas à interface INTERPRO (MULDER *et al.*, 2005, QUEVILLON *et al.*, 2005), obtendo-se maiores detalhes sobre as sequências.

Análises filogenéticas

O alinhamento global das sequências de aminoácido da ORF17 (KSA) com sequências obtidas no banco público de dados NCBI, foram conduzidas utilizando-se o programa ClustalX (LARKIN *et al.*, 2007). A análise filogenética foi realizada

empregando-se o pacote de programas PHYLIP, versão 3.6 (FELSENSTEIN, 2005). As árvores foram geradas a partir do alinhamento das sequências de aminoácidos através do programa PROTDIST e do método de distância NEIGHBOR, empregando os parâmetros Neighbour joining (NJ) e a matrix de substituição de aminoácidos JTT. Foram gerados 1000 agrupamentos de dados por meio do SEQBOOT, simulando-se a construção dos agrupamentos com “Bootstrapping” 1000. Uma árvore foi construída para cada réplica, e cada um dos “bootstraps” foram computados com o programa CONSENSE. A árvore obtida foi visualizada utilizando-se o programa NJplot (PERRIÈRE & GOUY, 1996). As árvores foram enraizadas pela entrada “F1 3-oxoacyl-(acyl-carrier-protein) synthase II” (Gi: 30254899 / *Bacillus anthracis*), a qual se constitui um membro da família “Fatty Acid Synthase F” (Fab-F), representando o grupo de cetosíntase de ácido graxo mais próximo filogeneticamente das PKS II de bactérias (JENKE-KODAMA & DITTMANN, 2009).

Comparação de “clusters”

O “cluster” metagenômico foi equiparado à outros agrupamentos biossintéticos da literatura, tanto de pigmento quanto relacionados à produção de antibióticos. Os esquemas priorizaram a composição gênica e a disposição química dos mesmos, e não estão em escala de tamanho relativa ao original.

2.4. RESULTADOS E DISCUSSÃO

Prospecção de genes para a PKS II

Foram prospectados 9.320 clones cosmidiais, em sistemas de 94 amostras compostas utilizando-se oligonucleotídeos iniciadores degenerados para uma região gênica conservada: *KS α* (METSÄ-KETELÄ *et al.*, 1999). Dentre as amostras compostas, três apresentaram amplificação por PCR próxima à região conservada

esperada: de 600 pares de bases (pb) de extensão (Fig. 16). A amplificação por PCR dos clones individuais resultou em 4 *amplicons*, dos quais 2 foram sequenciados (amostras B5P37 e C7P70), e um clone apresentou confirmação de homologia a região conservada: o clone B5P37. O sequenciamento completo do inserto cosmidial por biblioteca “*shotgun*” e pelo método adaptado de Sanger e a análise e montagem das sequências curtas lidas pelo sequenciador (“*reads*”) em um único projeto resultou em uma sequência nucleotídica consenso (“*consensus*”) de 22.988 pb, com conteúdo de bases guanina/citosina (G/C) correspondente a 72,6%. Esse conteúdo elevado de G/C também é encontrado no genoma de muitos micro-organismos produtores de compostos policetídeos (GUSTAFSSON *et al.*, 2004), dentre outros metabólitos secundários. A análise desta sequência nucleotídica por meio da ferramenta BlastN (ALTSCHUL *et al.*, 1990) permitiu verificar um resultado máximo de Cobertura da entrada/identidade de 30/79% com bactérias do gênero *Streptomyces* depositadas no NCBI, indicando uma estirpe não previamente sequenciada (79%), ao passo que o baixo valor referente a “cobertura de entrada” (30%) se justifica pelo fato da sequência de entrada ser menor que encontradas no banco de dados, limitada unicamente pelo tamanho do inserto acessado no clone da biblioteca.

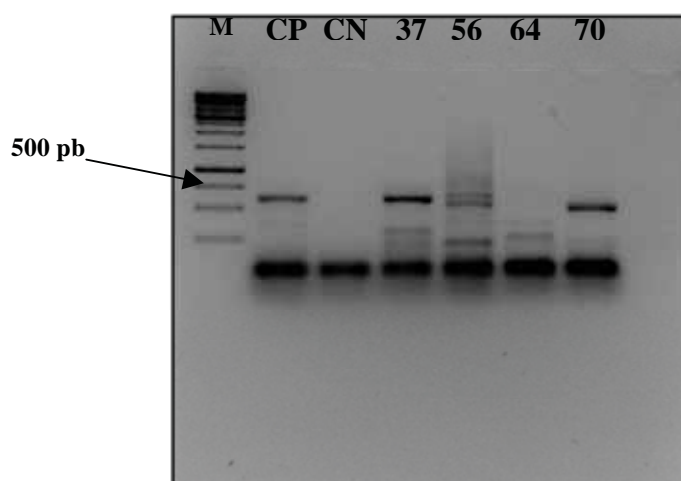


Figura 16. Amplificação das amostras compostas dos clones metagenômicos: 37, 56, 64 e 70, utilizando 100 ng de DNA para cada reação. **CN**: controle negativo; **CP**: controle positivo (*Streptomyces galilaeus*); **M**: Marcador de Tamanho Molecular 1 kb DNA “ladder” (Fermentas).

A Figura 17 ilustra o mapa de distribuição das ORFs obtidas (numeradas de 1 à 22) e dos conjuntos de domínios conservados (A à U) encontrados no “*consensus*”. A Tabela 1 trás uma descrição em maiores detalhes sobre os resultados de homologia BlastP e a Tabela A1 (Anexo 1) trás um resumo dos domínios encontrados no PRODOM para cada uma destas ORFs.

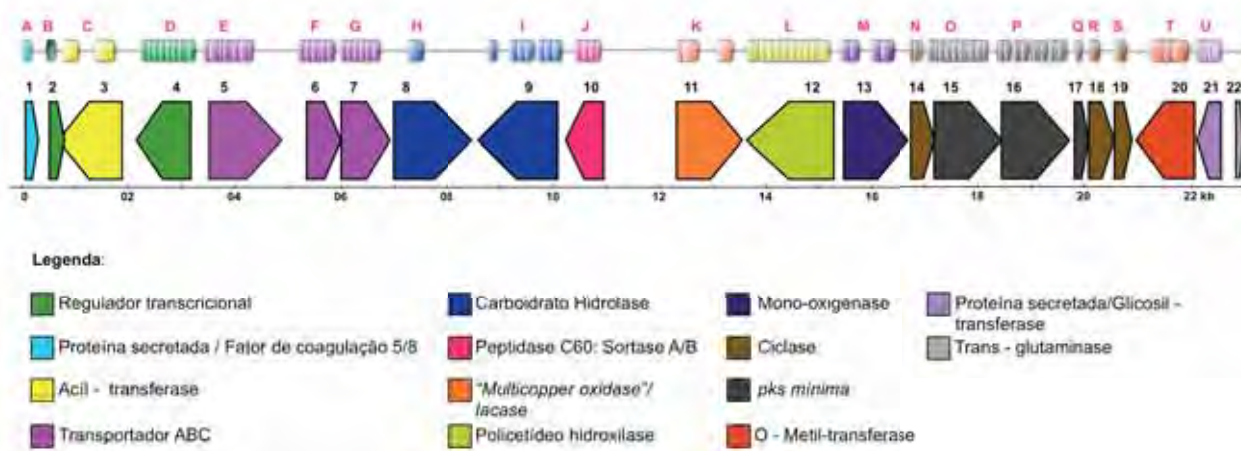


Figura 17. Mapa da distribuição de ORFs (1-22: Blastp) e Domínios conservados (A-U: PRODOM) obtidos para o “Cluster” metagenômico.

Inferências de função por homologia

ORFs para a *pks mínima*

Como pode ser observado na tabelas 1 (resultados do BlastP e no anexo A1 (resultados do PRODOM), as ORFs correspondentes a *pks mínima* são: ORF15 (*KSA*), ORF16 (*CLF*) e ORF17 (*ACP*). A **PKS Mínima** é um conjunto de enzimas compartilhado por todas as PKSII, e consiste nas subunidades da cetosintase: “Ceto-acil Sintase *alfa*” ($KS\alpha$ ou *KSA*) e “Fator Elongador de Cadeia” ($KS\beta$ ou *CLF*); e na Proteína Carreadora de Acíla (*ACP*). Na rota biossintética dos policetídeos aromáticos o início da cadeia (com acetato na maioria das vezes) e a sua alongação com malonato ou metil-malonato (dentre outros intermediários) são processadas pela PKS Mínima, através de uma ação integrada: a $KS\alpha$ (responsável pela ativação do sítio de condensação) e $KS\beta$ catalisam

as ligações tipo “Clasein” (ligação policetílica), auxiliados pela subunidade ACP, que serve como uma “âncora” à qual a cadeia crescente de policetídeo está covalentemente ligada por uma ligação tioéster.

Tabela 1- Resultados obtidos por homologia utilizando a ferramenta BlastP.

ORF	Length[#]	Similar (Blastp)	Micro-organismo /I.D.	S/I (%)
ORF1*	80	Proteína secretada	<i>Streptomyces sviveus</i> ZP_06921971.1	85/74
ORF2	94	Proteína Hipotética conservada	<i>Streptomyces pristinaespiralis</i> ZP_06908004.1	73/62
ORF3	398	Acil-transferase	<i>Streptomyces sviveus</i> ZP_06916349.1	59/51
ORF4	340	Família de reguladores LAC I	<i>Streptomyces hygrosopicus</i> ZP_05517570.1	87/78
ORF5	453	Transportador ABC de celobiose: proteína ligante de soluto	<i>Streptomyces pristinaespiralis</i> ZP_06908816.1	69/57
ORF6	214	Provável proteína permease transportadora de açúcar ABC	<i>Streptomyces sp. C</i> ZP_05504693.1	90/81
ORF7	305	Proteína integral de membrana transportadora de açúcar (ABC)	<i>Streptomyces sp. C</i> ZP_05504694.1	85/74
ORF8	493	Beta-galactosidase	<i>Streptomyces sp. e14</i> ZP_06708057.1	85/76
ORF9	498	Carboidrato hidrolase	<i>Streptomyces avermitilis</i> NP_826429.1	78/66
ORF10	237	Peptidase C60: sortase A e B	<i>Streptomyces sp. ACTE</i> ZP_06272147.1	66/53
ORF11	410	Multicopper oxidase tipe 2	<i>Conexibacter woesei</i> YP_003395347.1	47/35
ORF12	546	Policetideo hidroxilase / Mono-oxigenase FAD-ligante	<i>Streptomyces bingchenggensis</i> ADI05882.1 / <i>Streptomyces violaceusniger</i> Tu 4113/ZP_07608719.1	79/71 79/70
ORF13	408	Mono-oxigenase de biossíntese de Antibiótico	<i>Catenulispora acidiphila</i> DSM 44928/ YP_003111851.1	69/56
ORF14	149	Proteína TcmJ de síntese deTetracenomicina (ciclase I)	<i>Streptomyces hygrosopicus</i> ZP_07293932.1	89/79
ORF15	422	3-oxoacyl-ACP synthase I (Cetosintase alpha)	<i>Streptomyces ghanaensis</i> ZP_04689825.1	86/78
ORF16	424	3-oxoacyl-ACP synthase II, Chain Length Factor	<i>Streptomyces hygrosopicus</i> ZP_05513870.1	83/71
ORF17	86	Proteína Carreadora de açúcar (ACP)	<i>Streptomyces pristinaespiralis</i> ZP_06908839.1	70/58
ORF18	159	Ciclase I (TcmN ARO-CYC like)	<i>Streptomyces hygrosopicus</i> ZP_05513872.1	80/72
ORF19	109	Ciclase II	<i>Streptomyces hygrosopicus</i> ZP_05513873.1	87/81
ORF20	366	O-metiltransferase família 2	<i>Streptomyces sp. ACTE</i> ZP_06273109.1	85/73
ORF21	150	Proteína secretada/ Glicosil transferase	<i>Streptomyces hygrosopicus</i> ZP_05513877.1 <i>Cronobacter sakazakii</i> ABX51895.1	74/61 45/29
ORF22*	50	Similar à Transglutaminase	<i>Herbaspirillum Seropedicae</i> YP_003774067.1	54/42

* ORFs possivelmente interrompidas.

Genes de unidades enzimáticas de modelagem intermediária da cadeia

Algumas subunidades enzimáticas (Keto-redutases/KR, ciclases/CYC, aromatasas/Aro) cooperam com a PKS mínima para dirigir a modelagem da cadeia policetídica nascente (Weissman, 2009). O “cluster” apresentou ORFs homólogas a ciclases (CYC) e aromatasas/ciclases (ARO/CYC): ORFs 14, 18 e 19.

A ORF14 é homóloga à ciclases do tipo I de vários “clusters” de PKSII, tais como CYC I WhiE (proteína WhiE II-like) e de tetracenomicina (TcmJ) com I/S 89/79%. Foram também identificados 3 domínios enzimáticos no PRODOM para a ORF, correspondentes a CYC I de antibiótico para tetracenomicina e WhiESAI.

A ORF18 apresentou homologia pelo Blastp para uma Ciclase I/Aromatase (*TcmN ARO-CYC like*), e foi identificado o seguinte domínio para a região: “Ciclase Biosíntese de Policetídeo Antibiótico/Aromatase Síntase Oxidoreductase ORF Transferase Multifuncional”. As ARO/CYC constituem uma interessante família de subunidades enzimáticas que podem ocorrer em duas formas de arquitetura: um dos subgrupos é um di-domínio de 300 resíduos de aminoácidos, enquanto o outro é um monodomínio de 150 resíduos. Quando dividida ao meio, tanto a porção N-terminal quanto a C-terminal da enzima di-domínio são similares à sequência do monodomínio, mas a similaridade é muito maior na porção N-terminal. Quanto à forma de atuação, o di-domínio ARO/CYC catalisa a aromatização do primeiro anel de seis membros derivado do esqueleto do policetídeo em construção, após este ter sofrido uma redução no C-9. Já o monodomínio ARO/CYC, além de possuir a atividade aromatase, altera o padrão de dobramento do esqueleto policetídico nascente pela catalisação da condensação aldólica entre os carbonos da cadeia C-9/C-14, ao invés da condensação mais típica C-7/C-12 (ZAWADA e KHOSLA, 1997).

Os dois subgrupos diferem quanto ao substrato: O di-domínio ARO/CYC exibe uma hierarquia de especificidade para o tamanho de seus substratos de acordo com a PKS de origem: se a unidade enzimática provém de um “cluster” que produz moléculas de decacetídeos, ou seja formado por dez unidades extensoras de PKS (ex. griseusina) ela é capaz de catalisar a aromatização de substratos octacetídeos, nonacetídeos ou

deacetídeos, enquanto uma aroclase de octocetídeos (ex: actinorhodin) somente reconhece substratos octacetídeos. Ao que tudo indica, essa propriedade de maior controle sobre o substrato está relacionada à porção C-terminal diferenciada (ZAWADA e KHOSLA, 1997), e sua ausência no monodomínio confere a este um potencial de recombinação a vias de diversificados tamanhos de cadeia poliketídica.

A ORF18 apresenta um tamanho de 149 resíduos de aminoácidos, indicando um provável **monodomínio** ARO/CYC. A submissão ao PRODOM resultou em um domínio homólogo ao de uma “Óxido-redutase-Transferase Multifuncional”, o que também sugere sua participação na condensação aldólica característica dos monodomínios ARO/CYC.

Por sua vez, a ORF19 apresentou homologia à Ciclase II (CYC II), com S/I de 87/81%. As CYC II são uma classe de enzimas envolvidas na catálise do produto da síntese da ARO/CYC como substrato para formação de um segundo anel aromático na cadeia nascente. A identidade de CYC II da ORF19 também foi confirmada pela presença do domínio “Ciclase de Síntese de Policetídeo II/ ORF Síntase de JADI Biosíntese Similar a Ciclase de Antibiótico”, retornando uma identidade de 72% com a ciclase II homóloga da PKS do antibiótico jadamicina (uma anguciclina).

Não foi encontrada nenhuma ORF relacionada à enzima Cetoreductase (KR). As KR de PKSII catalisam uma reação de ceto-redução quimicamente idêntica à de KR das Ácido Graxo Síntases, embora com algumas diferenças determinantes de especificidade. Elas auxiliam algumas ciclases na síntese de anéis aromáticos, e da mesma forma que para a maioria das enzimas, excetuando-se a PKS mínima, sua presença no “cluster” não é obrigatória (WAWRIK *et al.*, 2005; TSAI e AMES, 2009).

Genes relacionados a modificação pós-síntese da cadeia poliketídica

Após a síntese e liberação da ACP, as cadeias poliketídicas podem, ainda, serem remodeladas por oxigenases e decoradas por: glicosil-, N-, O-, ou C- metil-transferases (dentre outras alcóis-transferases); amino-transferases; acil/acetil-

transferases; glicosil-transferases e quinases, além de muitas outras enzimas cujas contribuições metabólicas permanecem incertas (RIX *et al.*, 2002). A introdução de novos grupos funcionais altera tanto a reatividade como o perfil estereo-elétrico do substrato, e a riqueza das estruturas de policetídeos deriva destas modificações pós-síntese, responsáveis na maioria das vezes pela própria reatividade da molécula (WEISSMAN, 2009).

A ORF12 apresentou homologia BlastP com as enzimas “Policetídeo hidroxilase” e “Mono-oxigenase dependente de FAD”, além de 15 domínios para a região relacionados à “Mono-oxigenase Oxidoreductase Flavoproteína Ubiquinona FAD Hidroxilase”, indicando se tratar de uma enzima bifuncional mono-oxigenase/hidroxilase. A UrdM (*Streptomyces fradiae*) é um exemplo desse tipo de enzima: sua porção N-terminal é composta por uma flavina mono-oxigenase, e a região C-terminal exibe uma ação redutase. A UrdM é responsável pela formação do grupo **12b-hidroxiila angular**, o qual é indispensável para uma posterior adição de um desoxiaçúcar para síntese da anguciclina Urdamicina A (RIX *et al.*, 2002).

A ORF13 por sua vez obteve homologia à “Mono-oxigenase de Biossíntese de Antibiótico”. Uma vez que essas enzimas são classificadas de acordo com seus requerimentos de cofatores, o domínio “Mono-oxigenase de Biossíntese de Antibiótico Anthrone Oxigenase Transmembrana” chama atenção por destacar uma possível mono-oxigenase independente de cofator: as **anthrones oxigenases** são conhecidas como “mono-oxigenases internas”, ou seja, não necessitam de NADPH, NADH ou qualquer outro cofator (RIX *et al.*, 2002).

A ORF3 foi identificada como uma Acil-transferase, possuindo dois domínios “Acil-transferase Transferase TCP24 Precursor DBV8 Sinal” (ambos PD846045) identificados no PRODOM. A classe das acil-transferases: **dbv8 tcp**, são frequentemente empregadas em biossíntese combinatória, para a produção de antibióticos glico-peptídicos a partir de uma cadeia originalmente aglicana (sem açúcares). Nessas estratégias, normalmente se emprega primeiro uma glicosil-transferase, a qual adiciona amino-açúcares neutros na molécula, e em seguida utiliza-se uma enzima N-acil-transferase, tal como a **dbv8** ortóloga à **tcp**, para adição de

grupos acil, aumentando assim a diversidade química das moléculas recombinadas (SOSIO e DONADIO, 2006).

A ORF20 é homóloga à “O-metil-transferase/família 2”, confirmado por 4 domínios relacionados a essa enzima no PRODOM: dois especificados como “Ácida de Biossíntese Caféica de Lignina” e “OxyF”, respectivamente. As O-metil-transferases podem estar envolvidas tanto nos mecanismos de síntese de cadeias policetídicas, quanto na degradação de composto lignínico-celulósicos (HSIAO e KIRBY, 2008).

A ORF22 é pequena, com 50 resíduos de aminoácidos, e apresenta similaridade BlastP com S/I de 54/42% à uma superfamília de domínios trans-glutaminase (uma amino-transferase). No entanto esta ORF, localizada na porção final do *contig consensus*, aparentemente está interrompida em sua extremidade 3'. Esse tipo de enzima pode atuar na produção de radicais aminados, como os amino-açúcares constituintes de rifamicina, algumas antraciclínicas e macrolídeos (RIX *et al.*, 2002).

Genes relacionados a regulação transcricional e transporte de substâncias

Além de enzimas diretamente relacionadas a construção e modificação da cadeia policetídica, um “cluster” de PKS normalmente apresenta genes reguladores, tanto específicos como de atuação pleiotrópica (HUANG, *et al.*, 2005; HE *et al.*, 2008), e enzimas envolvidas na síntese de substratos não obtidos através das vias do metabolismo primário, como por exemplo alguns açúcares especiais. Outros componentes importantes do “cluster” são as enzimas e demais proteínas relacionadas a resistência ao antibiótico produzido, como os transportadores de membrana. Quanto a estes, muitos micro-organismos produtores de antibióticos possuem pelo menos um transportador ABC (“ATP-binding cassette”). Essa classe de transportadores de membrana atua no transporte ativo de açúcares, oligopeptídeos, aminoácidos e mesmo de drogas, o que lhes confere um importante papel nos mecanismos de resistência a antibióticos (MÉNDEZ e SALAS, 1998).

A ORF2 apresentou similaridade a reguladores transcricionais (S/I: 73/62%), e o

domínio conservado: “Regulador de Transcrição ligante de DNA, Família LuxR Lp_1789 Contendo Regulador”.

A ORF4 é similar a “Família de reguladores LAC I”, e tal homologia foi confirmada pela indicação de 8 domínios no PRODOM. Nos resultados obtidos, para a ORF4 no PFAM, foram localizadas duas famílias de domínios: a primeira com 45 resíduos de aminoácidos foi caracterizada como: “Proteínas regulatórias bacterianas, família Lacl” (CL0123); e a segunda de 259 resíduos de aminoácidos correspondeu a família: “Peripla_BP_1” (CL0144), o que confirma os resultados anteriores. A própria família Peripla_BP_1 trata-se de uma família que inclui reguladores transcricionais Lacl e proteínas periplasmáticas ligantes que atuam como receptores primários para quimiotaxia de muitos solutos carboidratos, remetendo a regulação do repressor Lac, uma vez que em atuação conjunta, a proteína se liga a um açúcar e dessa forma modifica a atividade ligante de DNA do domínio repressor. Por definição, Lacl é uma proteína regulatória de transcrição ligante de DNA por meio dos motifs helix-turn-helix (HTH) (Interpro, descrição IPR000843).

Para a classe de transportadores de membrana foram encontradas as seguintes ORFs : 5, 6 e 7. A ORF5 homóloga a “Transportador ABC de celobiose: proteína ligante de soluto” apresentou 9 domínios adjacentes no PRODOM, correspondentes a essa função. A ORF6, “Provável proteína permease ABC transportadora de açúcar” foi confirmada pela presença de 8 domínios no Prodom, e a ORF7, correspondente a “Celobiose transportador permease” também foi confirmada por 6 domínios relacionados ao gene (Tab.1-anexo).

Genes possivelmente envolvidos com PKS II e degradação de compostos lignínico-celulósicos e outros genes

Segundo HSIAO e KIRBY (2008), o gênero *Streptomyces* ocupa um importante papel no solo, atuando na reciclagem dos materiais de natureza lignínico-celulósica, e algumas enzimas envolvidas no processo parecem derivar de vias de policetídeos de

pigmento. A determinação de quais e quantas enzimas, precisamente, estariam envolvidas na rota metabólica de degradação de celulose e lignína (incluindo genes não relacionados a PKS) permanece obscura, mas é possível evidenciar a participação de: catalases/peroxidases; endoglucanases; celulases; proteínas ligantes de celulase; mono-oxigenases; dioxigenases; hemeoxigenases; tirosinases (monofenol/mono-oxigenases) e seus cofatores; celobiose-hidrolases; oxigenase; alcano-mono-oxigenases; e reguladores sensores de peróxido de hidrogênio, além de enzimas de metabolismo de celobiose. Algumas ainda podem também se relacionar novamente às vias de PKS de forma indireta, através do fornecimento de unidades precursoras para modificação da cadeia de policetídeos.

Uma vez que o “cluster” apresentou relação de homologia de algumas ORFs de açúcar-hidrolases e transportadores de celobiose, os resultados de homologia foram perscrutados na busca de possíveis genes envolvidos na degradação de lignína e celulose, paralelamente a busca por enzimas de síntese de policetídeos.

Conforme o exposto, dentre as ORFs já relatadas os candidatos a genes que podem estar envolvidos nesse processo, tem-se as ORF12 e a ORF13, que se relacionam respectivamente com as oxigenases: Policetídeo Hidroxilase/mono-oxigenase dependente de FAD e Policetídeo Mono-oxigenase. O papel de uma mono-oxigenase na degradação de compostos lignínico-celulósicos parece envolver a criação de radicais superóxidos que atuam atacando e desestabilizando estes compostos, favorecendo a atuação de outras enzimas (HSIAO e KIRBY, 2008). As ORFs homólogas à transportadores ABC de celobiose (ORF6, ORF8 e ORF9) também podem estar envolvidas no transporte de açúcares e enzimas secretadas para a via degradativa destes polissacarídeos, ao mesmo tempo em que o transporte de eventuais substâncias tóxicas possam subsidiar um mecanismo de resistência a um antibiótico eventualmente produzido.

A ORF8, a ORF9 e a ORF11 são homólogas a hidrolases de carboidrato. A ORF8 apresentou homologia BlastP com beta-galactosidase, e obteve o domínio “Hidrolase Beta-Glucosidase/Família 6-Phospho-Beta-Glicosidase” no PRODOM. O resultado da análise no PFAM indicou um domínio correspondente à família enzimática

das glicose-hidrolases tipo 1, uma representante de um vasto grupo de O-glicosil hidrolases que hidrolisa a ligação glicosídica entre dois ou mais açúcares, ou entre um carboidrato e um grupamento não açúcar. Fazem parte desta classe de carboidrases as enzimas: beta-glicosidases; beta-galactosidases; 6-phospho-beta-galactosidase; 6-phospho-beta-glicosidase; lactase-phlorizin hidrolase ; beta-mannosidase e mirosinase.

A ORF9 apresentou homologia BlastP com “Carboidrato hidrolase”, sendo os valores de S/I: 78/66%. O PFAM por sua vez indicou a presença do domínio “Celulose ligante” (CL0203), com localização de dois resíduos de triptofano que comporiam o sítio de ligação da celulose. O resultado obtido para o PRODOM apontou a localização domínios conservados relacionados a “quitinase secretada” e “Hidrolase quitinase transmembrana/Sinal Precursor Glicosidase metabolismo do carboidrato celulose”, indicando possível envolvimento no metabolismo dos polissacarídeos quitina e celulose.

A ORF11 apresentou homologia às “*Multicopper*” Oxidases (MCO). Sabe-se que as MCOs são uma família de enzimas que utilizam múltiplos átomos de cobre (Cu) como cofatores para o acoplamento da oxidação do substrato à redução de oxigênio em água. Os substratos de MCO incluem uma variedade de compostos orgânicos, bem como íons metálicos tais como o Fe^{2+} . As MCOs estão envolvidas em uma variedade de funções celulares, incluindo a homeostase ferro-cobre, oxidação siderófora, formação de pigmento e bio-polimerização. Uma das classes da MCOs são representadas por lacases. Estas possuem atividade benzenediol:oxigênio oxireductase, e são capazes de oxidar compostos fenólicos. Além desta atividade, as lacases estão envolvidas com a degradação de lignina (D'SOUZA-TICLO, 2008). Para essa última função a enzima exerce uma clivagem do tipo não específica, na região amorfa polímero de lignina. Diferente das peroxidases, seu cofator é o cobre e não um grupo heme, e não necessita de H_2O_2 como cosubstrato, mas sim oxigênio molecular (DICK *et al.*, 2008). Tendo em vista o exposto, a ORF11, como uma provável lacase, pode representar uma hidrolase capaz de degradar açúcares complexos atuando no meio da cadeia, o que a caracterizaria como uma possível endo-glucanase.

Tais resultados são compatíveis com a via de degradação de compostos celulósicos, sobretudo no caso de celulose amorfa, na qual é indispensável a presença

de três hidrolases: uma beta-glicosidase, uma celóbio-hidrolase (celulase) e uma endo-glucanase, possivelmente aqui representada por uma lacase (BETTIGA, GORWA-GRAUSLUND e HAHN-HAGERDAL, 2010).

A ORF21 possui como resultado de homologia indicações de “*Proteína secretada*” (S/I: 74/61%), algumas das quais descritas como ligantes de carboidratos. Para essa ORF foram identificadas 4 domínios no PRODOM, sendo que dois destes se referem a homologia a uma família de enterotoxinas de *Staphylococcus*. Outro domínio a define como uma proteína secretada. A região de maior escore (55%) se relaciona com um “Precursor de Sinal Repetitivo Paralelo Beta-Helix Kinase Epimerase de Superfície celular I 5.1.3”. Este domínio é descrito como envolvido na conversão do ácido beta-D-manurônico (M) ao ácido alfa-L-gulurônico (G) na superfície da célula. Devido a introdução quase que exclusivamente de blocos MG, tal enzima produziria um tipo de polímero que não possui capacidade de formar um gel. Um resultado mais baixo de homologia (S/I: 45/29%) com enzimas glicosil-transferase chamou a atenção pela estreita relação entre esta e a síntese de policetídeos, de forma que esta hipótese não foi descartada, deixando espaço para futuras investigações.

Foram também identificadas mais duas ORFs não diretamente relacionadas às anteriores: a ORF1 e a ORF10.

A ORF1 se localiza na extremidade inicial do “*contig consensus*”, e ao que tudo indica está interrompida. A ORF apresentou um domínio no PRODOM como uma hidrolase secretada. No entanto, outra possibilidade apresentada por resultados do BlastP é que possa ser uma proteína que funcione como um fator de coagulação conhecido como 5/8 em eucariotos, o qual apresenta um histórico de transferência horizontal para procariotos. A região estaria envolvida com mecanismos de intercâmbio de sequências entre diferentes organismos, compatível com as extensas transposições de vias de metabólitos secundários.

A ORF10 é homóloga a uma Peptidase C60: Sortase A e B. As Sortases fazem parte de uma classe de enzimas transpeptidases de membrana de bactérias gram-positivas, que cliva proteínas de superfície em sequências reconhecidas e catalisa uma reação de transpeptidação na qual o substrato proteína de superfície é covalentemente

ligado ao peptídeoglicano (NCBI-PDB). A região gênica apresentou 4 domínios conservados correspondentes no PRODOM, reforçando a homologia com Sortases C60 tipo B, com provável relação com sinais precursores de fimbrias na membrana plasmática.

Análise filogenética da KSA e Comparação entre “clusters”

A enzima KSA está relacionada com a escolha do precursor de extensão para a biossíntese dos antibióticos dessa família, de forma que diferenças em sua sequência podem refletir subclasses de diferentes moléculas (METSÄ-KETELÄ *et al.*, 1999). Sendo assim, foi feita uma análise filogenética entre a provável KSA metagenômica e outras sequências disponíveis no banco de dados, dentre as quais foram selecionadas algumas PKSII de pigmento de esporo e outras de antibióticos. As análises foram feitas com o auxílio dos programas ClustalX (LARKIN *et al.*, 2007) e Phylip (FELSENSTEIN, 2005), conforme descrito em procedimentos experimentais. O resultado da análise filogenética pode ser visto na Figura 18.

Pode-se notar que houve uma divisão dos grupos em dois clados, excetuando-se a KSA de uma Ácido Graxo Síntase (“*Fatty Acid Synthases*” -FAS) utilizada como grupo externo, e a KSA da *Streptomyces resistomycificus*. O primeiro clado (Fig.18-A) reuniu KSAs de conhecidas espécies produtoras de policetídeos aromáticos de antibióticos. Já o segundo se dividiu em dois ramos: um contendo KSAs relacionadas à produção de pigmentos de esporo (Fig.18-C), enquanto o outro reuniu três sequências (Fig. 18-B): a KSA metagenômica (“*Putative*” *KS-alpha Metagenômica*) e duas KSAs pertencentes a linhagens de *Streptomyces*: *S. hygroscópicus* ATCC 53653 (ZP_05513869.1) e *S. bingchenggensis* BCW-1 (ADI05885.1). Essas duas últimas são sequências depositadas no NCBI (banco **nr**) como prováveis KSA, sem definição de seu envolvimento com antibiótico ou pigmento. A *S. hygroscópicus* ATCC 53653 é conhecida pela produção do policetídeo Himastatina, o qual apresenta atividade antimicrobiana seletiva para bactérias gram-positivas, além de atividade anti-tumoral contra as células *P388 leukemia* e a *B16 melanoma* (LAM *et al.*, 1990). Por sua vez, *S.*

bingchenggensis teve seu genoma recentemente sequenciado, e a anotação preliminar de seus genes revelou a existência de, pelo menos, 23 “clusters” gênicos para policetídeos, peptídeos não ribossômicos e biossíntese de terpenos, sendo que a maior parte dos compostos que possam ser metabolizados por essas vias permanecem desconhecidos, a exceção de 4 vias para PKS I, dentre as quais tem-se a relacionada a produção do antibiótico nanchangmicina (WANG et al., 2010).

Desta forma, a identificação da KSA metagenômica (ORF15) por meio da análise filogenética de KSA não é clara, pois apesar de estar em um clado como grupo irmão de espécies de *Streptomyces* relacionadas a produção de pigmento de esporos, ela não entra dentro deste grupo. METSÄ-KETELÄ e colaboradores (1999) também obtiveram um grupo similar a partir de sequência de KSA amplificadas de organismos não identificados, sendo que os mesmos não foram atribuídos pelos autores como pertencentes ao grupo de pigmentos de esporo, e sim reunidos em uma nova categoria de grupo: o “clado 3” (METSÄ-KETELÄ et al., 1999: fig. 2.3, pg. 4). Por outro lado, a falta de caracterização precisa das KSAs para *S. hygrosópicus* ATCC 53653 (ZP_05513869.1) e *S. bingchenggensis* BCW-1 (ADI05885.1) não permitem a predição destas como oriundas de PKS de produção de pigmentos de esporos, ou envolvidas na produção de antibióticos, ou mesmo como diferenciadas em vias de degradação de lignina/celulose.

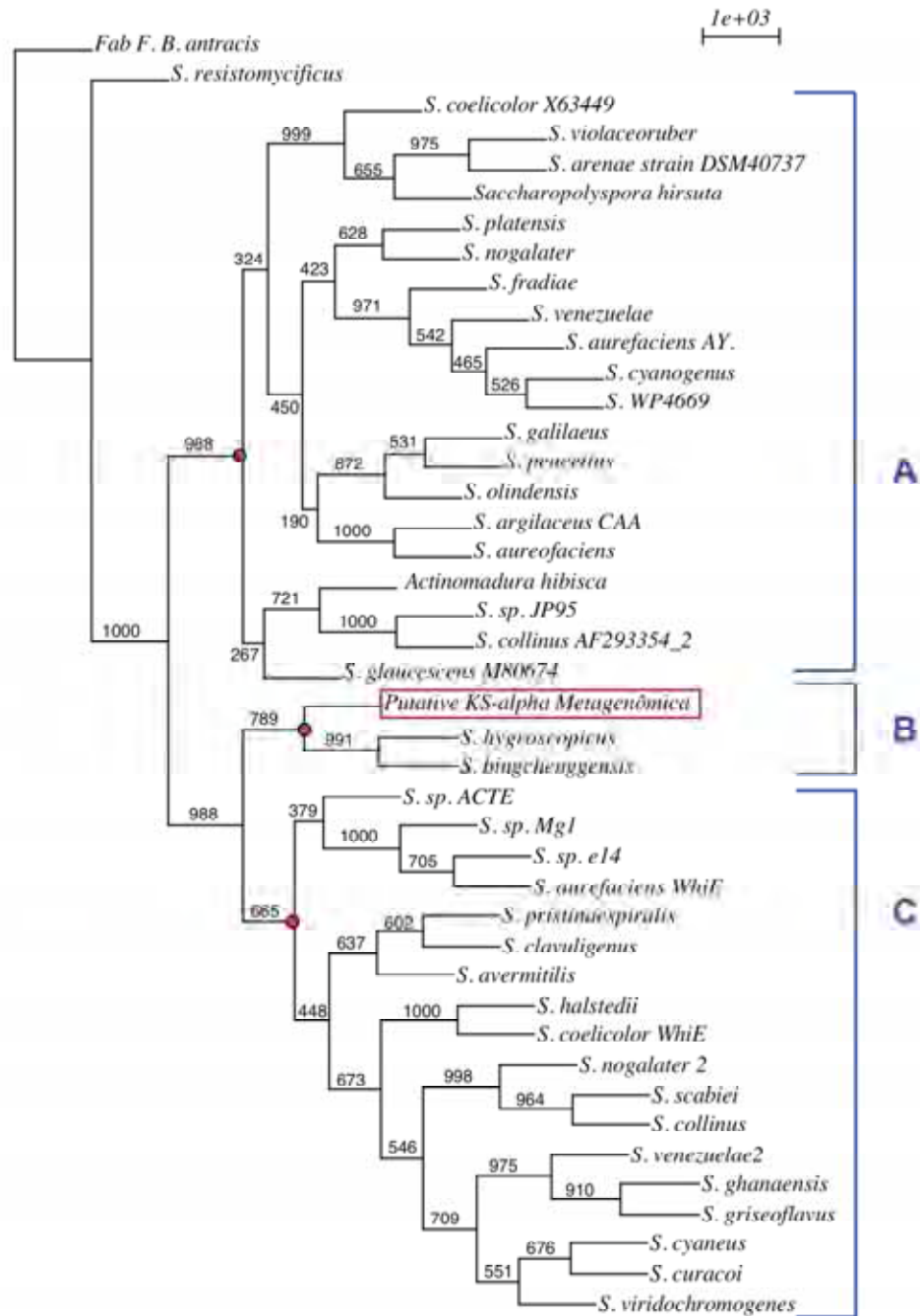


Figura 18. Comparação filogenética de genes para KSA de vários “clusters” de PKS disponíveis na literatura e a ORF metagenômica. A: PKS II de antibióticos; B: PKS II indefinidas; C: PKS II de pigmento de esporos. A árvore foi gerada utilizando o pacote de ferramenta de bioinformática Phylip (*Neighbour Joining*, matrix JTT, *Bootstrapping* 1000), e foi gerada para um alinhamento da sequência nucleotídica traduzidas para aminoácidos.

A Figura 19 é uma comparação entre o agrupamento biossintético obtido por metagenômica, e algumas PKSs II envolvidas na síntese de pigmentos ou na síntese de antibióticos. Pode-se notar uma maior semelhança dentro de cada grupo de “clusters” comparados, tanto na composição gênica quanto na disposição destes, e mesmo entre os genes dos dois grupos em menor escala.

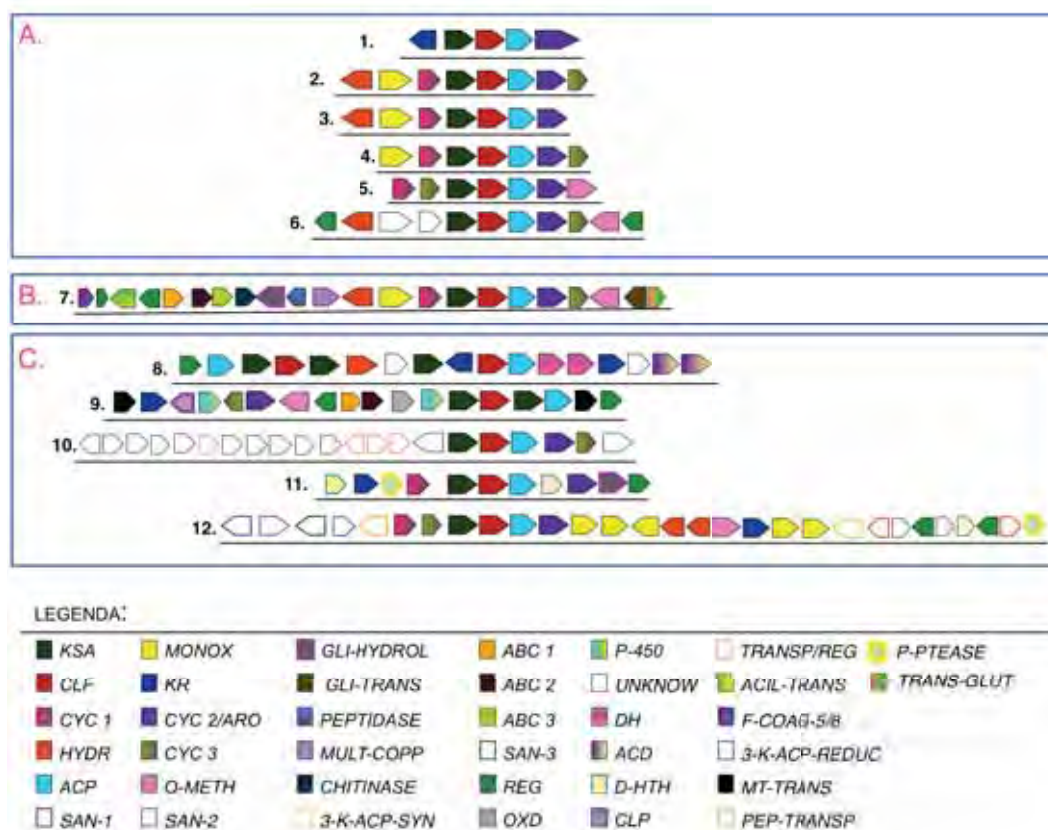


Figura 19. Análise comparativa entre o *contig* metagenômico **B5-37** e diversos “cluster” para PKS II. **A:** PKSs de pigmentos. **1:** *act* “cluster” (actinorrodin de pigmento) de esporo de *Streptomyces coelicolor* (YU *et al.* 1998) ; **2:** *whiE* “cluster” para pigmento de esporo de *S. coelicolor*; **3:** *sch* “cluster” para pigmento de esporo de *S. halstedii* (YU *et al.* 1998); **4:** *cur* “cluster” para pigmento de esporo de *S. curacoii* (YU *et al.* 1998); **5:** *tcm* “cluster” para tetracenomycin de pigmento de *S. glaucescens* (Yu *et al.* 1998); **6:** “cluster” de **Melanina** de *S. avermitilis* (OMURA *et al.*, 2001, fig. 2.A.3, p.12217). **B:** Agrupamento Gênico obtido para o *contig* metagenômico. **7:** “cluster” metagenômico **B5-37** **C:** Agrupamentos gênicos para antibióticos policetídeos. **8:** “cluster” do **Antibiótico 1** (OMURA *et al.*, 2001, fig. 3.C.1, pg. 12218) de *S. avermitilis*; **9:** “cluster” do **Antibiótico 2** (OMURA *et al.*, 2001, fig. 3.C.2 pg. 12218) de *S. avermitilis*; **10:** “cluster” do antibiótico **Actinorrodin** (BAERSON e RIMANDO, 2007); **11:** “cluster” do antibiótico **Antrachinona** de *Photorhabdus luminescens* (BRACHMANN *et al.* 2007); **12:** “cluster” do antibiótico **A-74528** (ZALETA-RIVERA *et al.*, 2010). **Observação:** as imagens não estão em escala de tamanho em relação ao “cluster” originais.

Essa característica é bastante comum para as PKSs, havendo um certo grau de compatibilidade entre enzimas provenientes de diferentes rotas, o que se constitui em uma excelente matéria-prima para a engenharia molecular e biossíntese combinatória (HOPWOOD, 1997).

Como pode ser observado, o agrupamento obtido apresenta um arranjo enzimático que se assemelha à disposição física de “clusters” relacionados aos pigmentos de esporo. Todavia, existe uma maior complexidade do agrupamento gênico, sobretudo de ORFs relacionadas a enzimas de modificações posteriores à síntese e de modelagem, o que parece ser mais condizente com a arquitetura de agrupamentos envolvidos na síntese de compostos antibióticos. Contudo, de forma similar aos dados obtidos para a filogenia da KS α , os resultados comparativos não permitem discriminar o “cluster” encontrado como de antibiótico ou de pigmentos. Ademais, a presença de homologia a genes de degradação de carboidratos pode indicar um “cluster” envolvido com a degradação de lignina, celulose e quitina, sendo que existem relatos na literatura sobre “clusters” derivados de PKS II que se diferenciaram e passaram a atuar neste processo (HSIAO e KIRBY, 2008).

2.5. Conclusões

Embora nem todas as vias de PKS estejam, originalmente, envolvidas com a síntese de compostos detentores de ação farmacológica, os genes de PKS possuem uma alta plasticidade, e podem ser intensamente combinados com as vias pré-existentes, podendo resultar em novas drogas, dentre outros metabólitos de interesse biotecnológicos.

A estratégia de caracterização proposta consistiu de uma abordagem simples, porém efetiva, para classificação das ORFs encontradas, o que é primordial para o rápido acesso à extensa quantidade de dados que foram colocados ao alcance da comunidade científica com o advento das estratégias metagenômicas. Os dados gerados poderão guiar experimentos de expressão e ensaios bioquímicos para testar e

explorar o potencial dos genes putativos caracterizados por homologia, uma vez que podem indicar possíveis substratos e interações enzimáticas. Além disso, em bibliotecas metagenômicas nem sempre o hospedeiro heterólogo escolhido (frequentemente *Escherichia coli*) é compatível com a expressão dos genes ambientais que abriga, podendo não haver compatibilidade entre seus códons de uso e os do organismo desconhecido. Levando-se em conta que o gênero *Streptomyces* e alguns outros *Actinomycetos* representam a maior fonte de metabólitos secundários (incluindo policetídeos), e possuem os códons de uso mais distintos e incompatíveis com a maioria dos organismos modelos conhecidos (GUSTAFSSON et al., 2004) o “screening” gênico se mostra uma alternativa altamente recomendável para explorar a riqueza metabólica representada pela Metagenômica.

O agrupamento gênico acessado por metagenômica corresponde à uma via de PKS II de uma estirpe não previamente sequenciada. A análise da KSA demonstra proximidade filogenética da via com os “clusters” envolvidos na produção de pigmentos de esporos, contudo o “cluster” metagenômico apresenta diferenciações suficientes para colocá-lo em um grupo a parte destes.

A despeito do grande número de trabalhos envolvendo estudos com PKS, ainda muito pouco se sabe do completo papel desempenhado por elas no ambiente em virtude da natureza extremamente diversa destas vias. Tal fato é evidenciado, cada vez mais, pela mudança de paradigmas envolvendo composição, cinética enzimática e utilização de substratos, dentre muitas outras atribuições, como a produção de sinais químicos desencadeadores de modificações fisiológicas, descoberta de novas moléculas químicas com intrincados padrões não usuais, e o recente relato de seu possível envolvimento com vias de degradação de celulose e lignina. A presença de carbohidrases intimamente próximas às enzimas de PKS II no inserto metagenômico analisado no presente trabalho desperta a curiosidade para abordagens futuras de investigação que contemplem também essa última possibilidade de atuação metabólica.

CAPÍTULO 3 – Predição estrutural *in silico* por homologia de uma PKS Mínima Metagenômica

Elisângela Soares Gomes, Viviane Schuch, Gabriel Padilla, Luciano Takeshi Kishi, Silvana Giuliatti, Eliana Gertrudes de Macedo Lemos

RESUMO - As Policetídeo Sintases Aromáticas (PKSII) são complexos enzimáticos biossintéticos envolvidos na síntese de antibióticos, pigmentos e sinais químicos responsáveis pela diferenciação fisiológica microbiana sob condições ambientais adversas. Neste trabalho foi realizada a modelagem estrutural por homologia das subunidades de uma enzima chave no processo: a KS (KSA/CLF), a qual foi previamente obtida por prospecção em bibliotecas metagenômicas. Os modelos gerados foram avaliados quanto à acurácia empregando os softwares: Verify3D, Qmean, Procheck, Rampage e Swiss-Model. As estruturas foram comparadas às resolvidas por métodos experimentais, e foi possível verificar uma conservação em determinadas regiões (tais como as responsáveis pela arquitetura do sítio ativo), enquanto outras apresentaram estrutura secundária peculiar. A análise de interação entre as subunidades (“docking”) permitiu simular a geometria gerada entre as predições de subunidades monoméricas. A presente metodologia para caracterização estrutural de novas enzimas de PKSII *in silico* representou uma abordagem de identificação acessível e rápida, gerando dados que podem auxiliar futuras pesquisas em biossíntese combinatória para desenho racional de novas drogas, permitindo selecionar enzimas de arquitetura inédita acessadas por metagenômica que apresentem compatibilidade com outras enzimas da via a ser estudada.

Palavras - chave: PKSII, “Docking”, Metagenômica, Predição 3D *in silico*

3.2. Introdução

As Policetídeo Sintases Aromáticas (PKS II) representam um importante grupo de vias biossintéticas presentes em micro-organismos e plantas, e estão envolvidas na síntese de antibióticos policetídeos (principal fonte de formulações farmacêuticas disponíveis), pigmentos e sinais químicos responsáveis pela diferenciação funcional diante de condições desfavoráveis no meio. Dentre as enzimas do “cluster”, a KS (formada pelas subunidades KSA e CLF) representa uma enzima-chave nos processos de escolha de substratos e controle sobre o tamanho da cadeia química dos compostos gerados (METSÄ-KETELÄ *et al.*, 1999).

Ainda são muito escassos os dados referentes à estrutura tridimensional e à organização na forma de complexo enzimático das PKS II, o que sem dúvida limita as estratégias de desenho racional de policetídeos aromáticos. Contudo, dados sobre estruturas 3D de algumas enzimas envolvidas na via permitem o emprego de estratégias que visem esclarecer como se dão as interações interprotéicas (CASTALDO *et al.*, 2008). Um importante marco para o início dos estudos das interações entre as proteínas da via foi a resolução por cristalografia de raio-X da proteína KS em sua forma heterodimérica (KSA/CLF) de *Streptomyces coelicolor* (KEATINGE-CLAY *et al.*, 2003), permitindo conhecer a geometria da cavidade formada pelas subunidades responsáveis pela catálise das reações de condensação descarboxilativa que configura a ligação policetídea. Juntas, a KSA, a CLF, e a ACP formam um complexo enzimático que ditará a natureza do substrato e o comprimento de cadeia, denominado de PKS Mínima. A atração não covalente entre elas é responsável mesmo pela aquisição da sua forma funcional e a arquitetura do sítio ativo. Além destas, muitas vezes o complexo enzimático depende de uma interação com outras enzimas para se tornar ativo. As ciclases, por exemplo, parecem atuar de forma decisiva para a estabilidade do complexo enzimático, prevenindo a eliminação de cadeias muito curtas de policetídeos no sítio ativo, garantido um melhor rendimento da via (CASTALDO *et al.*, 2008).

3.3. Material e Métodos

Análises de modelagem estrutural

Algumas ORFs foram submetidas a modelagem estrutural por homologia (KELLEY e STERNBERG, 2009). Para tanto procedeu-se o alinhamento das mesmas a sequências de proteínas cuja a estrutura terciária fora previamente resolvida por métodos experimentais (RMN e Cristalografia de Raio-X), disponíveis no Banco de dados de Proteínas PDB (NCBI), pelo emprego do software **Phyre** (<<http://www.sbg.bio.ic.ac.uk/phyre/>>; KELLEY e STERNBERG, 2009). Os moldes mais similares (acima de 30% de identidade com as ORFs) foram empregues para a construção de modelos de estruturas 3D pelo uso da Modelagem automática por homologia, utilizando o programa do Swiss-model (ARNOLD et al. 2006). Os modelos foram submetidos às análises de avaliação da compatibilidade à estruturas reais (“assessing”) empregando os softwares disponíveis na interface Swiss-model: (<http://swissmodel.expasy.org/>): Qmean, Procheck, Anolea. Os modelos foram também submetidos aos servidores Prosa-web (<https://prosa.services.came.sbg.ac.at/prosa.php>), Verify3D: (http://nihserver.mbi.ucla.edu/Verify_3D/), Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).

As análises de simulação entre subunidades enzimáticas foram realizadas utilizando o programa GRAMM-X (TOVCHIGRECHKO & VAKSER, 2006).

3.4. Resultados e Discussão

Modelagem 3D in silico e análises de compatibilidade com modelos reais (“assessing”)

Avanços no desenvolvimento de programas de modelagem de estrutura tais como “PSI-Blast”, “O Modelo Oculto de Markov” e “Algoritmos de sobreposição perfil-à-perfil” permitem uma predição e modelagem relativamente acurada de novas proteínas

com menos de 20% de identidade com uma proteína de estrutura conhecida, demonstrando o quão poderosas e promissoras estas abordagens são para o estudo de novas rotas biossintéticas (KELLEY e STERNBERG, 2009; BALTZ, 2006). Ao contrário dos modelos por ***Ab initio***, os quais são limitados a predição de estruturas pequenas (menos de 100 resíduos de aminoácidos) com base nas características químicas de sua estrutura primária, estes realizam a **modelagem por homologia** (modelagem com base em moldes homólogos) e “**fold-recognition**” (modelagem por alinhamento de estruturas: o chamado “*threading*”). As aplicações da predição da estrutura terciária das proteínas abrangem a formação de hipótese funcional para proteínas hipotéticas, proporcionam sinais de fase em cristalografia, permitem a seleção de sítios para a mutagênese e o desenho racional de novas drogas (KELLEY e STERNBERG, 2009).

Embora as predições para sequências com similaridade abaixo de 20% sejam possíveis, a confiabilidade da modelagem é diretamente proporcional a identidade residual da proteína alvo e do modelo. Valores acima de 40% de identidade retornam modelos preditivos de alta confiabilidade, ideais para estratégias de “docking” (predição de ligantes em sítios ativos, como drogas ou mesmo substratos). Valores entre 30-40%, são empregados para elaboração de inferências evolutivas. Abaixo de 30%, os modelos apresentam muitas falhas e extensões de gaps e erros de alinhamento, e passam a ser modelos preditivos limitados, mas ainda passíveis de utilização para verificação geral de padrões aproximados de estruturas desconhecidas (MARTI-RENO *et al.*, 2002). Para melhorar a possibilidade de predição em estruturas para as quais existam uma quantidade limitada de moldes, o cuidado com a escolha destes, a acurácia no alinhamento local, e o emprego de alinhamentos múltiplos é uma estratégia bastante útil.

Após confecção dos modelos estes devem ser checados para verificar o padrão de confiabilidade em relação às estruturas reais, sobretudo na região do sítio ativo protéico, verificando a correspondência entre o modelo e o molde. O modelo criado sobre as restrições químico-espaciais do molde deve ter boa qualidade estéreo-química, refletida pelos padrões adequados de comprimento de ligação, ligação peptídica e planaridade 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. Tais características podem ser obtidas pelo uso do programas PROCHECK (LASKOWSKI et al., 1993). Adicionalmente, os padrões de características espaciais compilados de estruturas conhecidas de alta resolução para o modelo podem ser testados por programas que trabalham com métodos baseados no perfil 3D e potenciais estatísticos de força, tais como o VERIFY3D ([WWW.VERIFY3D: http://nihserver.mbi.ucla.edu/Verify_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/); LÜTHY et al., 1992; BOWIE et al., 1992), PROSA WEB (WIEDERSTEIN e SIPPL, 2007). Outros programas, como o QMEAN (BENKERT et al., 2009), permitem uma avaliação simultânea do modelo como um todo, e de cada um dos aminoácidos tomados localmente, retornando resultados com base nos parâmetros de escores combinados. Tomados em conjunto estes valores irão indicar quais as informações biológicas podem ser inferidas a partir de um dado modelo, garantido que as extrapolações feitas a partir dos resultados assumam proporções confiáveis, limitados à capacidade preditiva do modelo. No presente trabalho a modelagem das ORFs foi feita empregando as ferramentas disponíveis na interface Swiss-model (ARNOLD et al. 2006). Para escolha dos moldes foram realizados alinhamentos das ORFs à sequências já elucidadas pelo método clássico, empregando o programa PHYRE (KELLEY & STERNBERG, 2009), buscando selecionar preferencialmente moldes que apresentaram alinhamentos que resultassem identidade igual ou maior à 30%. Após a síntese dos modelos estes foram avaliados e comparados aos moldes, conforme o exposto na Tabela 2.

Tabela 2 - Avaliação estéreo-química molde/modelo

3D/L	Id. (%) ^a	Regiões obtidas no gráfico de Ramachandran (Procheck) ^b				Escore 3D Profile*		
		Reg. mais favorável (%)	Reg. permitida(%)	Reg. generosamente permitida(%)	Reg. não permitida(%)	Total	Ideal	S _{ideal}
1TQYA (394)	57	87,5	11,3	0,6	0,6	155,98	180,22	0,86
ORF15 (416)		85,6	13,6	0,3	0,6	183,07	190,36	0,96
1TQYH (402)	47	87,2	11,5	0,9	0,3	192,39	183,90	1,04
ORF16 (402)		86,2	12,2	0,6	0,9	163,75	183,90	0,89
1NQ4A (95)	33	74,1	21,0	2,5	2,5	34,32	42,96	0,79
ORF17 (85)		73,6	18,1	2,8	5,6	25,24	38,40	0,65

^a: resultado para o molde mais similar (foram empregados alinhamentos múltiplos), Id.: Identidade. ^b: Gráfico de Ramachandran (PROCHECK) baseado nos ângulos de torção Φ e Ψ (C-alpha e N e C-alpha e C-peptídico). *Dados calculados a partir da análise no programa Verify3 (http://nihserver.mbi.ucla.edu/Verify_3D/) Escore Total: fornecido. Escore ideal: $\text{Exp}^{-0,83+1,008 \times \ln(L)}$, e S_{ideal} , que equivale a compatibilidade da sequência com sua estrutura 3D, e é obtido por meio da divisão entre o escore total e escore Ideal, sendo que valores ideais estão acima de 0,45 S_{ideal} . L.: número de resíduos de aminoácidos (DA SILVEIRA *et al.*, 2005).

Um modelo estrutural de alta precisão (tanto modelagem experimental quanto *in silico*) deve ter em torno de 90% de seus aminoácidos localizados na região mais favorável do gráfico de **Ramachandran (PROCHECK)**. No entanto, assume-se que acima de 85% obtenha-se modelos com um bom índice de confiança (DA SILVEIRA *et al.*, 2005). Como pode ser observado na Tabela 2, que relaciona os resultados obtidos tanto para o modelo da ORF, quanto para um dos moldes correspondentes utilizados pelo PHYRE, somente a ORF17 e seu molde não atingiu esse valor, sendo também a de menor identidade com os moldes (30%). A ORF15, por sua vez, foi a que apresentou maior valor de identidade (57%), o que possibilitou uma estrutura de alta resolução, conforme pode ser observado pelo Z-Score, que é comparável à resoluções alcançadas por uso de Cristalografia de Raio - X (Fig.20-B, gráfico à direita: região em azul-claro). Embora seu índice **Ramachandran (PROCHECK)** tenha retornado 0,6% de aminoácidos na região não permitida, pode se observar no gráfico obtido para o **Ramachandran Atualizado (RAMPAGE: LOVELL *et al.*, 2003)** que os dois resíduos são de glicina (Fig.20-B, centro), de forma que essa região não lhe é realmente proibitiva, em vista de sua cadeia lateral curta (apenas um H). Esse resultado reflete também a alta conservação destas proteínas.

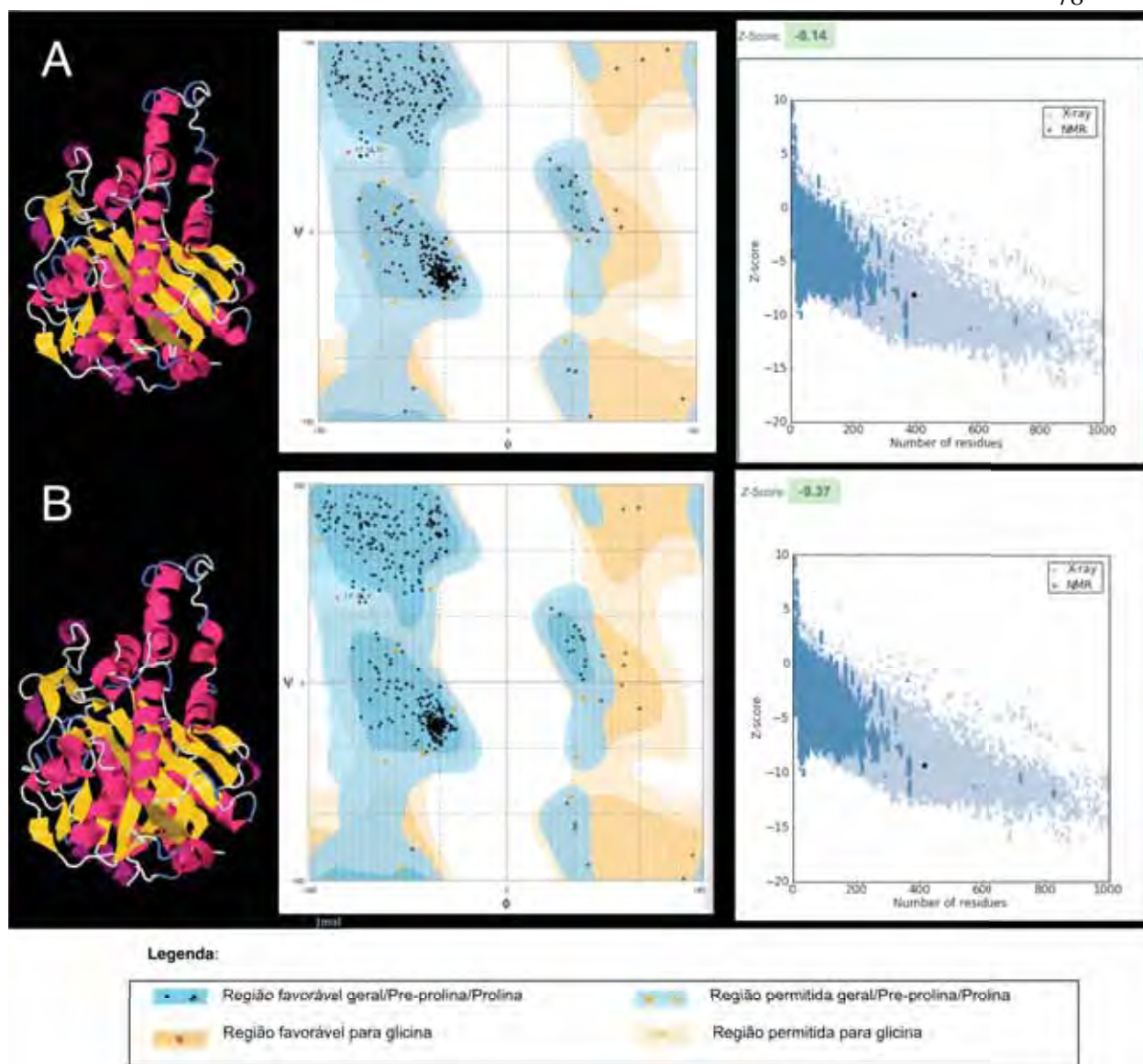


Figura 20. Predição funcional para o modelo e a ORF15. Centro: Ramachandran (Rampage) Baseado nos ângulos de torção C-alpha e C-beta. Direita: Z-score (Prosa Web).

O modelo predito poderia ainda ser utilizado em estudos tipo “docking” enzima-substrato, ideais para predição de substrato, o que seria altamente desejável uma vez que as KSAs estão envolvidas com a escolha do substrato de extensão (malonato, metil- ou etil- malonato dentre outros substratos derivados) e essa especificidade permite por vezes diferenciar vias de classes distintas de antibiótico (METSÄ-KETELÄ *et al.*, 1999). Abaixo (Fig. 21), tem-se uma análise comparativa obtida através da

análise utilizando o programa Procheck para a estrutura secundária da ORF15 e do molde 1TQYA.

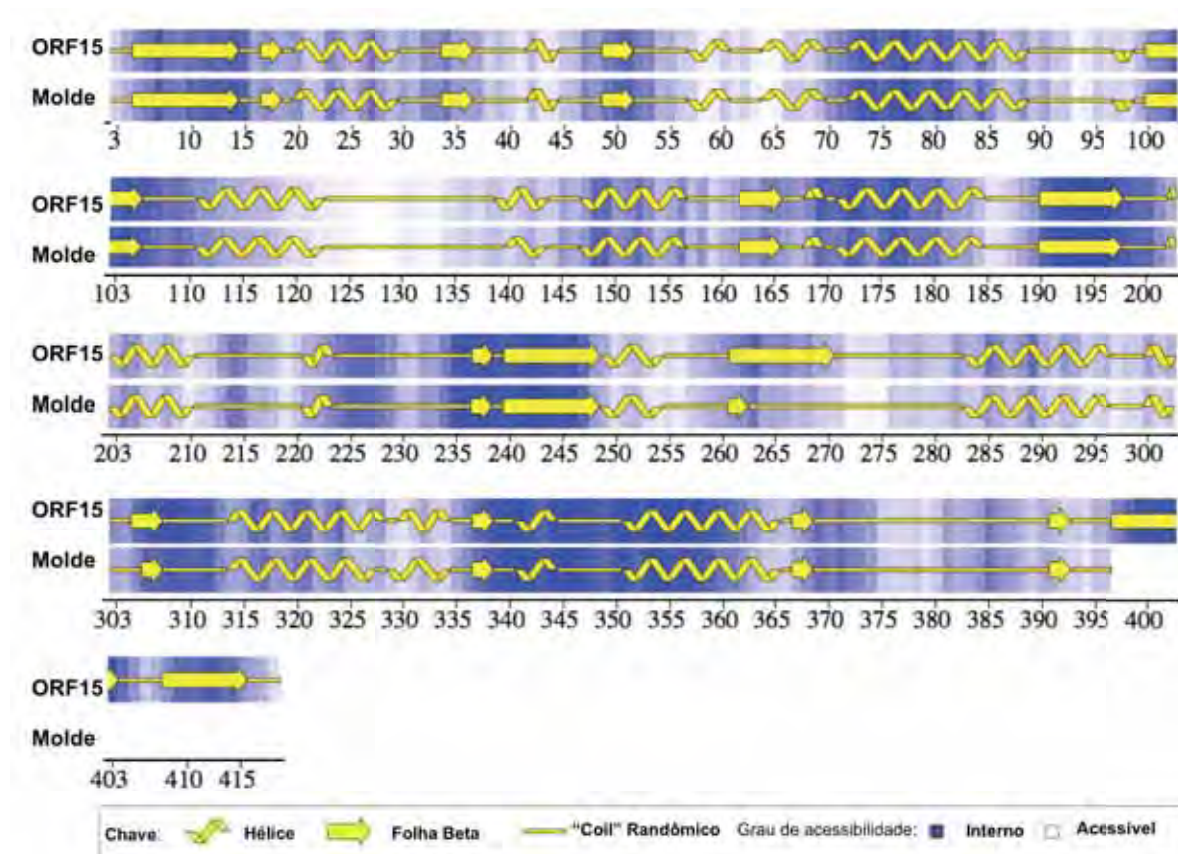


Figura 21. Comparação da estrutura 2D entre a ORF15 e o molde 1TQYA.

Como pode ser observado na figura, a ORF15 difere da estrutura do molde, sendo maior do que este e apresentando duas estruturas de folha beta à mais (região peptídica: 397-418). Além disso, apresenta diferenças no comprimento de estruturas de folhas betas nas regiões peptídicas 260-270 e 304-307, onde é possível evidenciar folhas beta maiores na estrutura da ORF. Essas diferenças na estrutura secundária irão afetar a estrutura protéica total, podendo inferir diretamente sobre a característica funcional da mesma, interferindo no potencial redox e na afinidade por substratos.

A Figura 22 apresenta os resultados de escore Qmean para a ORF15 e o molde.

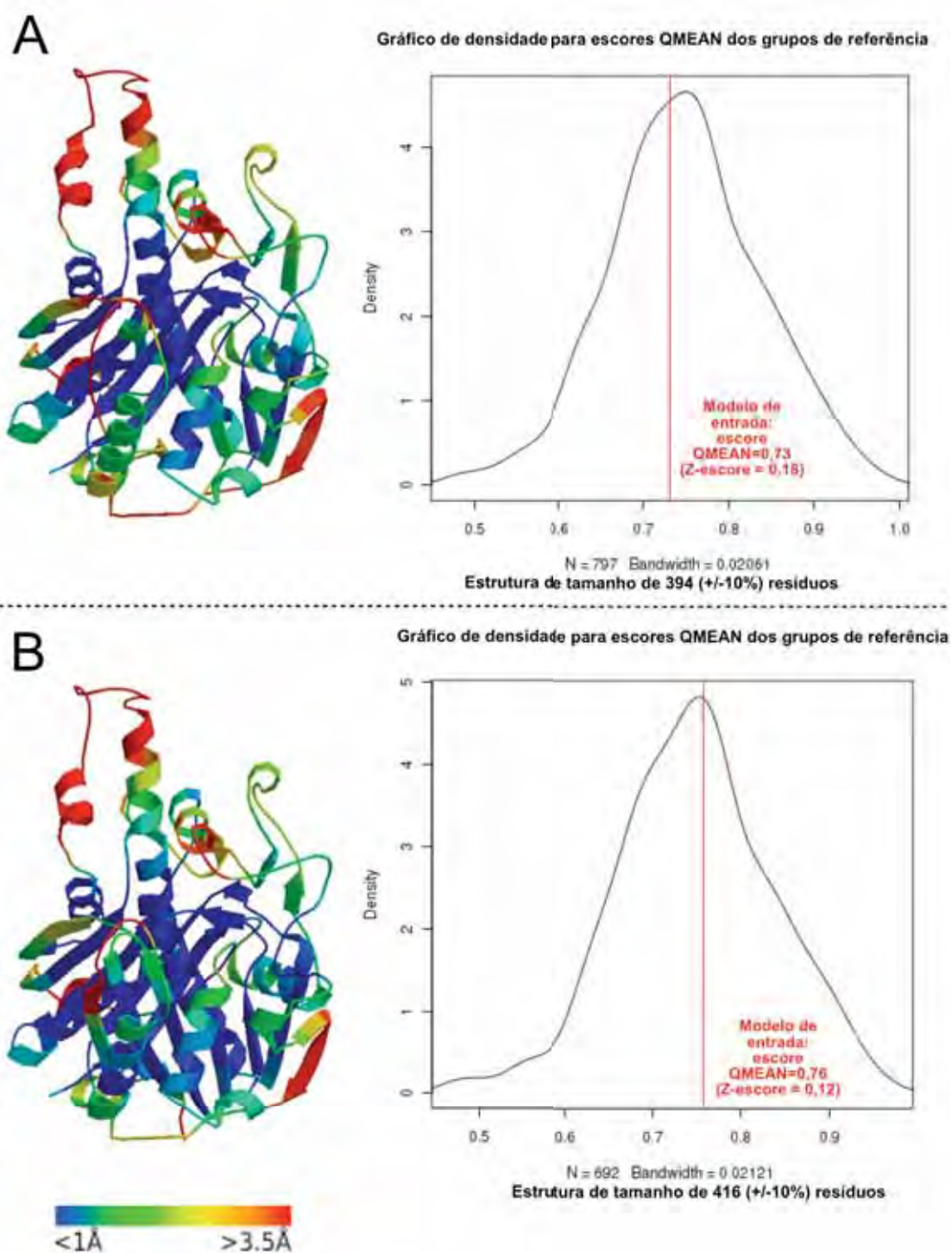


Figura 22. Avaliação da qualidade da estrutura 3D por meio do Qmean. A: Molde 1TQYA; B: ORF15.

A análise de escores fornecidas pelo programa Qmean permite avaliar o quanto um determinado modelo de estrutura 3D assume a configuração mais próxima da esperada para o padrão combinado de uma determinada composição em resíduos de aminoácidos, levando-se em conta as interações entre estes e os diferentes ambientes químicos que possam assumir. A análise é empregada para avaliar tanto modelos 3D obtidos por metodologias clássicas (RMN e Cristalografia de Raio-X) quanto modelos preditos *in silico*. A escala de cores do Qmean indica o quão próximo o modelo está do ideal, sendo as regiões indicadas em azul de alta qualidade de predição (desvios inferiores a 1 Å) e as regiões em vermelho as de mais baixo poder preditivo (desvios superiores a 3,5 Å). A ORF15 apresentou um padrão estrutural que se assemelha em qualidade ao obtido para a estrutura molde, com a maioria dos resíduos ocupando uma conformação ideal. Conforme pode ser visto na Figura 22, nos gráficos à direita de cada modelo, o padrão geral de escores avaliados pelo Qmean ficaram dentro da normalidade para ambas estruturas.

A estrutura do molde empregado para modelagem das estruturas das ORFs 15 e 16 (1TQY) é na verdade uma estrutura octaédrica (Fig. 23), formada por quatro monômeros idênticos, sendo cada um destes constituído por um dímero KSA/CLF. Contudo, a formação octaédrica não é encontrada para todas as KS, podendo ser estas constituídas até por apenas por um único dímero KSA/CLF.

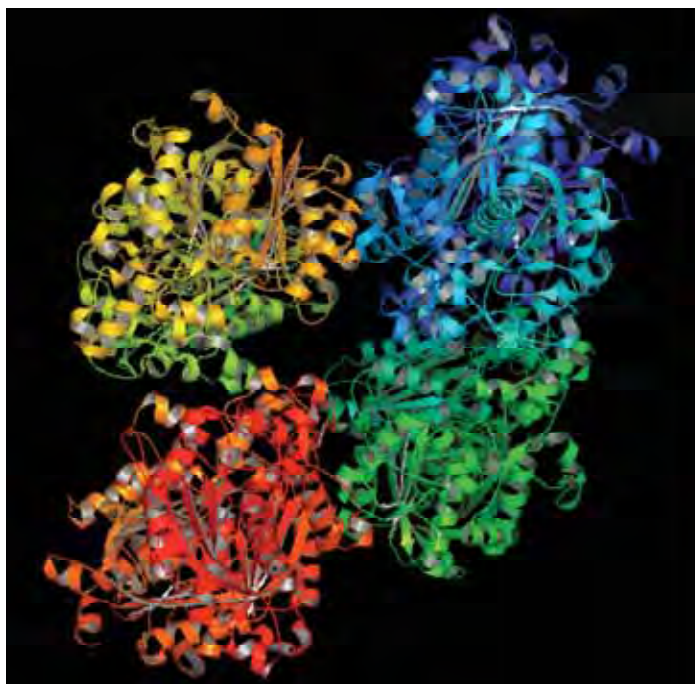


Figura 23. Modelo 1TQY na forma octomérica, formada por 4 dímeros CLF/ACP (PDB).

Embora ainda não se possa afirmar o número de dímeros que farão parte da estrutura final da PKS mínima metagenômica é possível visualizar a interação das mesmas ao nível de cada monômero, uma vez que a principal interação entre as cadeias se dá por interligação entre as KSA e CLF (CASTALDO et al., 2008). Foi feita uma análise de simulação de contato proteína-proteína (“docking”) entre as ORFs 15 (KSA) e 16 (CLF), buscando analisar a disposição das duas subunidades da KS. Na simulação buscou-se reproduzir uma situação em que fosse possível visualizar a enzima na forma monomérica, sendo que este arranjo foi equiparado com um monômero isolado da estrutura 1TQY. O resultado da simulação pode ser visto na Figura 24.

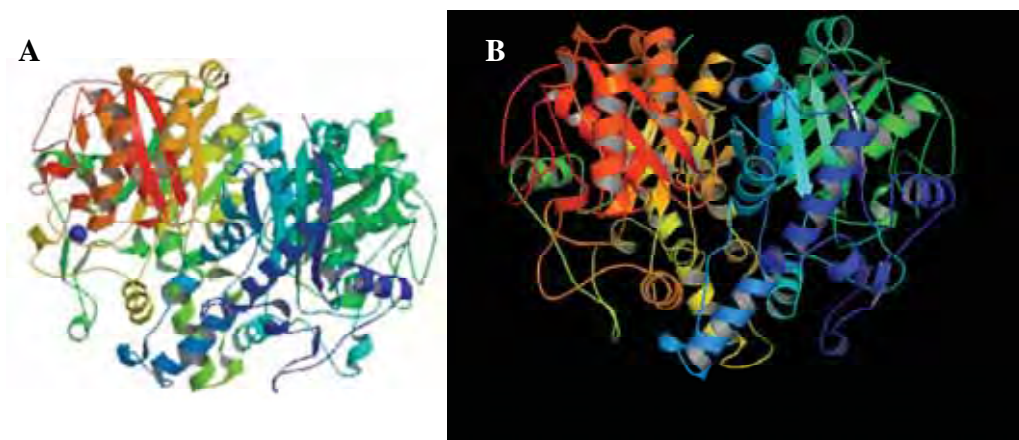


Figura 24. Comparação entre a estrutura de um dos dímeros 1TQY (KSA+CLF) e os resultados obtidos para o “docking” entre a subunidade metagenômica (22. B) KSA (laranja/vermelho) e CLF (Azul e verde).

Pode-se notar que existe uma semelhança muito grande na disposição geométrica assumida pelas estruturas metagenômicas submetidas a uma simulação de contato ao acaso, em que não foram previamente direcionados os resíduos para a interação (software GRAMM-X), e a estrutura encontrada para um dos monômeros do molde.

Além da KS (KSA/CLF), a PKS Mínima conta com a presença da “Proteína Carreadora de Acila” (ACP). Abaixo (Fig. 25), tem-se a comparação entre as estruturas das subunidades ACP metagenômica (ORF17) e o molde de maior identidade, onde é possível verificar uma grande semelhança entre as estruturas. Contudo, é possível evidenciar que a ORF17 apresenta uma estrutura em “coil” bem menor em relação ao molde (estruturas da extremidade, em vermelho), e uma substituição de uma pequena alfa-hélice localizada na porção mediana em amarelo no molde, por uma prolongação do “loop” na estrutura metagenômica (também em amarelo).

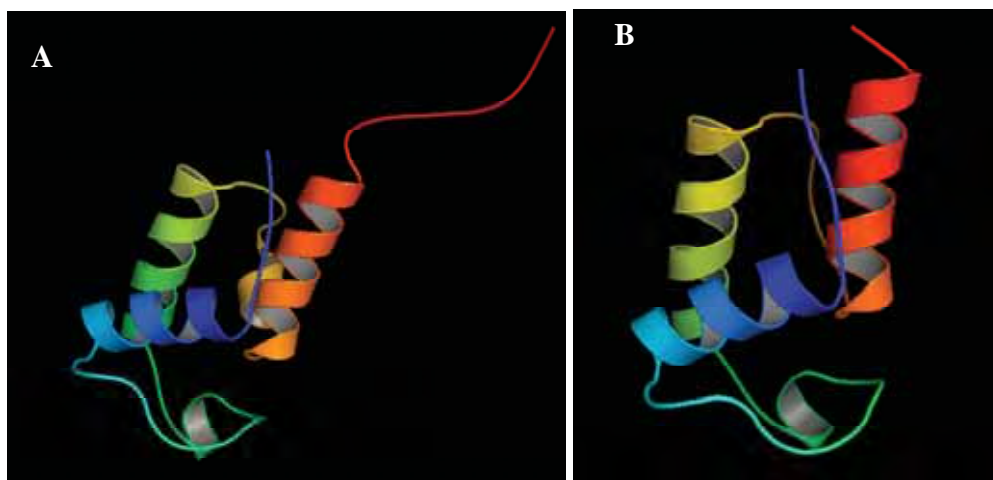


Figura 25. Comparação entre as estruturas da ACP 1NQ4A e o modelo preditivo para a ACP metagenômica. A: estrutura do molde (RMN). B: modelo de estrutura obtido para a ACP metagenômica.

Os resultados de “assessing” (avaliação da compatibilidade entre o modelo e o ideal), tanto para a ACP predita (metagenômica) e ACP molde, determinada pelo método clássico (RMN), representaram resultados de baixa qualidade, como pode ser observado na Tabela 2. Contudo, os valores de S_{ideal} acima de 0,45 indicam não haver erros grosseiros na estrutura. Além disso, o modelo gerado por experimentação clássica foi previamente validado por seus autores (FINDLOW et al., 2003).

Com o objetivo de verificar a compatibilidade do modelo da KSA metagenômica com outras CLFs obtidas na literatura por metodologia experimental, foi feito um “docking” proteína-proteína entre eles utilizando o programa GRAMM-X (TOVCHIGRECHKO & VAKSER, 2006). Para tanto, empregou-se os dois moldes de maior identidade obtidos por submissão da ORF16 (CLF metagenômica) ao Phyre: o molde 1tqyH (47% de identidade), relativo à KSA de “actinorhodin polyketide putative beta-ketoacyl”, e o molde 1j3nA (34% de identidade), relativo à “crystal structure of 3-oxoacyl-(acyl-carrier protein) 2 synthase II from *Thermus thermophilus* hb8”, relacionada a FAS, e o molde 1e5mA, referente a uma CLF de cadeia única, do organismo *Synechocystis* sp. (MOTCHE et al, 2001). A Figura 26 representa os dados obtidos.

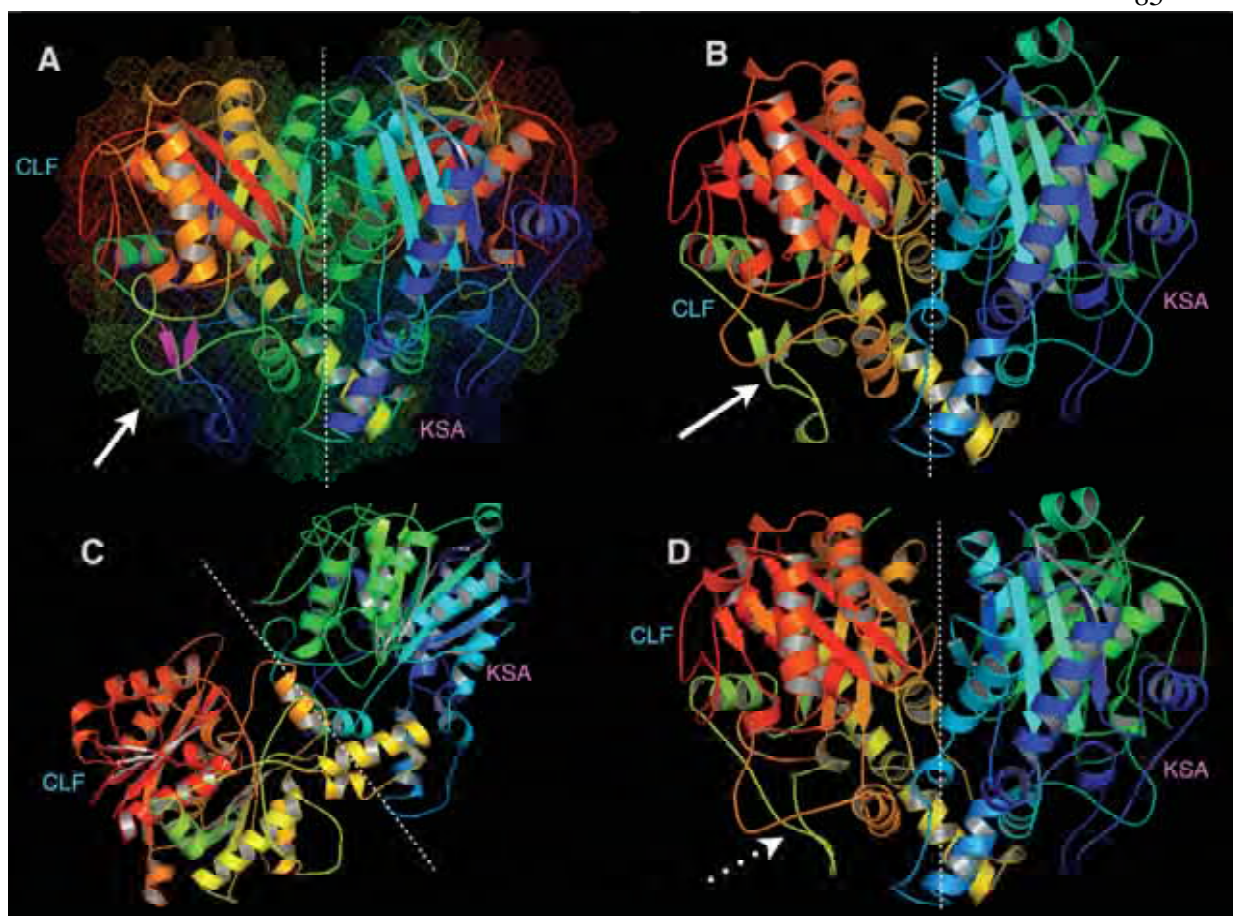


Figura 26. Análise de compatibilidade entre KSA Metagenômica contra a CLF Metagenômica (A), CLF 1tqyH (B), CLF 1j3nA (C) e CLF 1e5mA (D), obtidas por simulação de “Docking” proteína-proteína (GRAMM-X). Setas contínuas: detalhe para as folhas-beta curtas e antiparalelas, Seta pontilhada: ausências destas estruturas.

Pelas imagens da Figura 26, percebe-se que as simulações B e D, foram as que retornaram uma predição de contato entre proteína que refletiu melhor a situação encontrada para o modelo metagenômico (A). A CLF utilizada para o modelo B (1tqyH) é encontrada em sua origem ligada a KSA 1tqyA, formando um dímero que se combina com mais 3 dímeros idênticos, formando uma estrutura octaédrica (Fig. 23). Já a CLF utilizada em D, representa uma subunidade da cetosintase que se liga a uma KSA, também formando um dímero, mas que se apresenta em cópia única. Uma das diferenças entre as duas CLFs pode ser visualizada pelo detalhe indicado nas setas,

mostrando a ausência no modelo D das estruturas de folhas-beta anti-paralelas indicadas para a CLF metagenômica (A) e a CLF 1e5mA (D).

A situação de simulação C retornou um modelo totalmente diferente ao arranjo dos modelos metagenômicos. Essa situação era esperada, uma vez que as CLF 1J3nA é na verdade pertencente à FAS e não à KSA, indicando homologia por parálogos. Este resultado permite visualizar uma situação em que não existiria compatibilidade entre as enzimas candidatas a um ensaio de biossíntese combinatória, representando uma forma rápida de diagnóstico prévio das chances de se obter sucesso na recombinação enzimática, uma vez que existem expectativas de utilização da enzima de condensação de KS de ácido graxo sintases para a produção de novas drogas, explorando o parentesco entre as vias de PKS e FAS (MOTCH et al, 2001).

FISCHIBACH e colaboradores (2008) inferiram que a diversificação e propagação das PKSs entre vários grupos de micro-organismos se dá pela **evolução coletiva** dos genes. Os autores indicam que os “clusters” microbianos são os elementos gênicos que mais rapidamente sofrem evolução, sujeitos tanto a eventos comuns de transferências horizontais quanto a transferência vertical mais frequente dado o curto tempo de geração destes organismos. Em virtude do rápido diagnóstico de mudança fenotípica por mudanças de vias, o grupo também assume importância como modelo de estudo da evolução conjunta de genes.

Um fator decisivo para a propagação das vias de policetídeos é o fato de que muitas vezes estas pequenas moléculas irão conferir vantagens adaptativas ao seu hospedeiro, e muitas vezes são propagados por plasmídeos. O mecanismo que mantém a evolução dos policetídeos é o elevado custo de manutenção destas vias pelos organismos que e as possuem: a partir do momento que a mesma não confere vantagem adaptativa para o organismo hospedeiro, torna o seu tempo de vida evolutivo bastante curto (FISCHIBACH et al., 2008). Essa seleção negativa (aumento do gasto metabólico), aliada à seleção positiva por vantagens adaptativas conferidas por vias bem sucedidas, criam um crivo muito eficaz para a inovação de moléculas bioativas, ampliando muito o leque de possibilidades para o campo das pesquisas de PKS.

Além dessa forte seleção natural, o emprego da biossíntese combinatória pode ampliar em muitas ordens de magnitude a capacidade de recuperação de policetídeos. Essa abordagem se beneficia do fato de que embora cada rota bioquímica normalmente resulte em um tipo específico de moléculas, irão existir dentro de uma mesma família de PKS enzimas comuns a todas as rotas, derivadas de genes homólogos e detentores de regiões conservadas, ainda que não sejam idênticas entre si.

Esta conservação de propriedades dentre as enzimas homólogas permite certa compatibilidade entre grupos enzimáticos de “clusters” diferentes. No caso de antibióticos como os da Família PKS II, em que a produção da molécula final é resultado da interação de várias enzimas pertencentes a um “cluster”, a biossíntese combinatória permite criar moléculas através da recombinação de módulos enzimáticos de organismos produtores já conhecidos, bem como combinações de módulos conhecidos com genes novos obtidos pela prospecção em metagenômica, o que basicamente pode ser entendido como uma estratégia para acelerar os processos evolutivos, visando “competir” com a evolução natural dos antibióticos, buscando a síntese de novas moléculas bioativas para a farmacologia (BALTZ, 2006).

Para se ter uma idéia de quanto promissora é essa abordagem, deve-se considerar a imensa quantidade de complexos peptídeos, policetídeos e outros arcabouços químicos conhecidos, amplamente diversificados pela presença e disposição de estruturas tais como: glicosilações, metilações, halogenizações, grupos acílas e hidroxilações, e que se desenvolveram ao longo de bilhões de anos em actinomicetos e fungos (BALTZ, 2006).

Como se não bastasse, os estudos no “core” mais conservado da PKSII, a **pks mínima**, tem mostrado muitas diferenças entre enzimas provenientes dos mais diversificados “clusters”, que vão desde escolha específica para unidades precursoras, até a compatibilidade com outras enzimas do processo, tais como ciclases, cetoreductases e aromatasas.

SHEN (1999) descreveu um interessante trabalho em que organismos recombinantes foram produzidos a partir da introdução de uma *pks mínima* derivada de PKSII de esporo de pigmento, resultando em uma variedade de novos esqueletos

carbônicos muito superior a encontrada em ensaios que empregaram a *pks mínima* de produtores naturais de antibióticos.

4.5. Conclusões

A metagenômica é sem dúvida alguma uma ferramenta poderosa para a recuperação de novas vias enzimáticas, e esse trabalho permitiu avançar mais alguns passos na análise preditiva de função de proteínas por homologia, lançando luz sobre hipóteses de dobramento tridimensional destas moléculas. O presente estudo permitiu caracterizar enzimas metagenômicas putativas de uma forma rápida, totalmente *in silico*, o que permitiu confirmar a homologia entre as ORFs estudadas, detectando diversificações no padrão de estrutura 3D e 2D, bem como possibilitando o uso de “docking” para restituir subunidades enzimáticas da KS, permitindo simular a estrutura quaternária da KS, em um primeiro passo para o estudo de eventuais ligantes e substratos. Foi possível simular também situações de recombinação entre as subunidade KSA metagenômica e CLFs da literatura, podendo-se observar variações entre as diferentes situações de recombinações.

CAPÍTULO 4 – Análise *in silico* de uma O-metil-transferase de interesse biotecnológico acessada por metagenômica

Elisângela Soares Gomes, Viviane Schuch, Silvana Giuliatti, Luciano Takeshi Kishi, Gabriel Padilla, Eliana Gertrudes de Macedo Lemos.

RESUMO - Policetídeos são uma importante classe de metabólitos secundários complexos que apresentam os mais variados padrões químicos, graças a uma atuação conjunta de diversas enzimas nos processos de sua biossíntese, e constituem a maioria dos antibióticos conhecidos, dentre outros fármacos. A exceção do “core” biossintético de PKS mínima (KSA, CLF e ACP) e algumas enzimas relacionadas à direta modificação e formação das moléculas químicas durante a síntese (CYC, ARO, KR), após a liberação da cadeia policetídica da ACP, as combinações das demais enzimas das inúmeras vias de policetídeos são imprevisíveis. Uma importante classe catalítica de modificação das cadeias policetídicas são as transferases, principalmente as alcóois-transferases, que podem ser dos tipos: glicosil-, N-, O-, ou C- Metil-transferases. As transferências de tais grupos funcionais são importantes por serem capazes de alterar tanto a reatividade como o perfil estereo-elétrico do substrato, e a riqueza das estruturas de policetídeos deriva destas modificações pós-síntese, responsáveis na maioria das vezes pela própria reatividade da molécula. Neste trabalho, descreve-se a caracterização por homologia e identificação de domínios enzimáticos de uma O-metil-transferase recuperada por metagenômica, bem como a predição estrutural enzimática por homologia e validação do modelo gerado.

Palavras-chaves: O-metil-transferase, policetídeos, estrutura 3D, PKS

4.2. INTRODUÇÃO

Os micro-organismos praticamente sustentam a vida no planeta atuando em todos os ciclos biogeoquímicos, no controle de populações, na reciclagem de matéria orgânicas. Importantes patógenos de plantas, animais e mesmo outros micro-organismos, constituem também a maior fonte de moléculas bioativas para a farmacologia. Dentre estes compostos, os mais abundantes são os policetídeos, responsáveis por um faturamento anual superior aos U\$ 20 bilhões (WEISSMAN, 2009).

Contudo, contrastando com os eficientes mecanismos dos micropatógenos para aquisição de resistências aos fármacos em circulação, o processo de obtenção de novas formulações farmacológicas é bastante lento. Nos últimos 30 anos, apenas duas classes de antibióticos inéditas foram lançadas no mercado: a oxazolidinona linezolida e o lipopeptídeo cíclico daptomicina. Como agravante, a telavancina (um glicopeptídeo), a retapamulina (uma pleuromutilina) e a tigeciclina (última geração de tetraciclinas) foram os únicos princípios ativos de antibióticos admitidos como moléculas totalmente novas pela “*US Food and Drug Administration*” (FDA) lançados no mercado nos últimos cinco anos (HAMAD, 2010). Além disso, a intensa prospecção de novos fármacos policetídeos (iniciada em 1950, com a descoberta da estreptomicina) empregando métodos tradicionais de isolamento e cultivo de micro-organismos não é mais tão eficaz devido, principalmente, à alta taxa de redescoberta de antibióticos já conhecidos, que chega a 99.9% (ZAEHNER & FIELDLER, 1995).

Atendendo a crescente demanda da farmacologia frente à plasticidade de adaptação dos patógenos, os policetídeos têm sido alvos da engenharia molecular desde 1985 para desenho de compostos bioativos “sob medida”: mais eficientes, menos tóxicos ou inéditos (HOPWOOD *et al.*, 1985; RIX *et al.*, 2002). Para tanto, os pesquisadores tem lançado mão da biossíntese combinatória, que consiste na criação de novas rotas biossintéticas graças à recombinação de enzimas provenientes de diferentes complexos enzimáticos.

A utilização da Metagenômica como abordagem de busca destes genes de

interesse biotecnológico amplia em grande escala a perspectiva de sucesso destas estratégias, uma vez que abrange uma diversidade não cerceada pela possibilidade ou não de cultivo laboratorial (a qual seria de apenas 1%). A ampla diversidade microbiana do solo constitui-se de um enorme “pool” genético e biológico que pode ser explorado para a descoberta de novos genes, vias metabólicas inteiras e seus produtos (COWAN, 2005; BALTZ, 2006).

Na rota biossintética dos policetídeos aromáticos o início da cadeia e a sua elongação com malonato ou metil-malonato (dentre outros intermediários) são processadas pela PKS Mínima, através de uma ação integrada da $KS\alpha$ e $KS\beta$ para síntese de ligações tipo “*clasein*”, conjuntamente com a subunidade ACP, que serve como uma “âncora” à qual a cadeia crescente de policetídeo está covalentemente ligada por ligação tioéster. Algumas subunidades enzimáticas (cetoredutases-KR, ciclases-CYC, aromatasas-Aro) cooperam com a PKS mínima para dirigir a modelagem da cadeia policetídica nascente.

Após a síntese e liberação da ACP, as cadeias policetídicas podem, ainda, serem remodeladas por oxigenases e decoradas por inúmeras transferases (RIX *et al.*, 2002). A introdução de novos grupos funcionais altera tanto a reatividade como o perfil estereo-elétrico do substrato, e a riqueza das estruturas de policetídeos deriva destas modificações pós-síntese, responsáveis na maioria das vezes pela própria reatividade da molécula (WEISSMAN, 2009).

Avanços no desenvolvimento de programas de modelagem de estrutura tais como “PSI-Blast”, o Modelo Oculto de Markov (“Hidden Markov Model”) e Algoritmos de sobreposição perfil-à-perfil (“Profile–profile matching algorithms”) permitem uma predição e modelagem relativamente acurada de novas proteínas com menos de 20% de identidade com uma proteína de estrutura conhecida, demonstrando o quão poderosas e promissoras estas abordagens são para o estudo de novas rotas biossintéticas (KELLEY e STERNBERG, 2009; BALTZ, 2006). Ao contrário dos modelos por **Ab initio**, os quais são limitados a predição de estruturas pequenas (menos de 100 resíduos de aminoácidos) com base nas características químicas de sua estrutura primária, estes softwares permitem realizar a **modelagem por homologia** e a

modelagem por alinhamento de estruturas (“*fold-recognition*” ou “*threading*”). As aplicações da predição da estrutura terciária das proteínas abrangem a formação de hipótese funcional para proteínas hipotéticas, proporcionam sinais de fase em cristalografia, permitem a seleção de sítios para a mutagênese, além de permitir o desenho racional de novas drogas (KELLEY e STERNBERG, 2009).

Embora as predições para sequências com similaridade abaixo de 20% sejam possíveis, a confiabilidade da modelagem é diretamente proporcional à identidade residual da proteína alvo e do modelo. Valores acima de 40% de identidade retornam modelos preditivos de alta confiabilidade, ideais para estratégias de predição de ligantes em sítio ativos, como drogas ou mesmo substratos (“*docking*”). Valores entre 30-40% são empregados para elaboração de inferências evolutivas. Abaixo de 30%, os modelos apresentam muitas falhas e extensões de gaps e erros de alinhamento, e passam a ser modelos preditivos limitados, mas ainda passíveis de utilização para verificação geral de padrões aproximados de estruturas desconhecidas (MARTIRENOM *et al.*, 2002). Para melhorar a possibilidade de predição em estruturas para as quais existam uma quantidade limitada de moldes, o cuidado com a escolha destes, a acurácia no alinhamento local, e emprego de alinhamentos múltiplos é uma estratégia bastante útil.

Após confecção dos modelos estes devem ser checados para verificar o padrão de confiabilidade em relação às estruturas reais, sobretudo na região do sítio ativo protéico, verificando a correspondência entre o modelo e o molde. O modelo criado sobre as restrições químico-espaciais do molde deve ter boa qualidade estéreo-química, refletida pelos padrões adequados de comprimento de ligação, ligação peptídica e planaridade 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. Tais características podem ser obtidas pelo uso do programas PROCHECK (LASKOWSKI *et al.*, 1993). Adicionalmente, os padrões de características espaciais do modelo copilados de estruturas conhecidas de alta resolução podem ser testados por programas que trabalham com métodos baseados no perfil 3D e potenciais estatísticos de força, tais como o VERIFY3D (WWW.VERIFY3D:

http://nihserver.mbi.ucla.edu/Verify_3D/; LÜTHY et al., 1992; BOWIE et al., 1991), PROSA WEB (WIEDERSTEIN e SIPPL, 2007). Outros programas como o QMEAN (BENKERT et al., 2009) permitem uma avaliação simultânea do modelo como um todo, e de cada um dos aminoácidos tomados localmente, retornando resultados com base nos parâmetros de escores combinados.

Tomados em conjunto estes valores irão indicar quais as informações biológicas podem ser inferidas a partir de um dado modelo, garantido que as extrapolações feitas a partir dos resultados assumam proporções confiáveis, limitados à capacidade preditiva do modelo. Nesse trabalho foi feita a análise *in silico* de uma O-Metiltransferase obtida através do sequenciamento de um fragmento de DNA metagenômico. A enzima foi caracterizada segundo a presença de domínios enzimáticos, e foi submetida à predição de sua estrutura 3D visando subsidiar iniciativas futuras de emprego da mesma em estratégias de biossíntese combinatória.

4.3. MATERIAL E MÉTODOS

Biblioteca

A biblioteca cosmidial foi construída conforme descrito em um trabalho anterior (SCHUCH, *et al.* 2009) usando o “*pWEB-TNC Cosmid Cloning Kit*” (Epicentre Biotechnologies Madison, WI, USA) de acordo com as instruções do Fabricante, e representa clones de fragmentos de DNA ambiental extraído de um solo sob serrapilheira de bosques de eucaliptos não sujeitos a tratos culturais, ou revolvimento de substrato por atividade humana, por cerca de 40 anos (SILVEIRA et al. 2006).

Prospecção de genes de PKS

A prospecção de genes para PKS II foi conduzida na biblioteca metagenômica previamente construída, por meio de amplificação por PCR utilizando oligonucleotídeos

degenerados complementares a uma região altamente conservada do “cluster”: a KS α (METSÄ-KETELÄ *et al.*, 1999), sendo que para a padronização da reação e controle do padrão de amplificação positiva foram utilizados DNA genômico de *Streptomyces chartreuses* e *S. rimosus* e *S. galilaeus*, extraídos pelo método de MARMUR (1961).

Como estratégia de otimização as amostras dos clones foram reunidas em 94 “pools” de 96 clones cada antes da extração de DNA dos cosmídios pelo método de lise alcalina (SAMBROOK e RUSSEL, 2001). A amostras compostas de DNA foram então empregues como molde das reações de amplificação por PCR. Somente os pools que apresentaram amplificação positiva foram submetidos a uma extração individual de DNA cosmidial, seguido de amplificação por PCR para identificar quais os clones de amplificação positiva, seguido de sequenciamento do cosmídeo.

Subclonagem, sequenciamento do cosmídeo escolhido e análises de bioinformática

O sequenciamento do clone cosmidial selecionado foi conduzido utilizando-se a técnica de biblioteca “shotgun”. O clone foi submetido a nebulização por 12 segundos à pressão com nitrogênio gasoso de 3 Kg/cm². o DNA nebulizado foi submetido a uma eletroforese preparativa, na qual foram selecionados fragmentos entre 1,5 à 2 kb de extensão. Estes foram clonados em vetores *pUC19DNA/SmaI* (Tools), obtendo-se 10 placas de 96 clones cada. O DNA plasmidial foi obtido através do método de extração por lise alcalina (SAMBROOK e RUSSEL, 2001). Os clones da sub-biblioteca foram submetidos à PCR de sequenciamento utilizando o *DYEnamic ET Terminator* (Amersham Biosciences, São Paulo, Brasil) com os oligonucleotídeos M13F e M13R.

A eletroforese foi conduzida em sequenciador capilar automático ABI 3700 (Applied Biosystems, São Paulo, Brasil) de acordo com as instruções do fabricante. Os dados brutos dos eletroferogramas foram analisados utilizando o pacote de programas Phred/Phrap (EWING *et al.*, 1988; EWING e GREEN, 1998) e Consed (GORDON *et al.*, 1998) para gerar as sequências FASTA e avaliar a qualidade das sequências

individuais (“reads”) e das resultantes da montagem (“assembly”). Valores de Phred acima de 20 foram considerados de boa qualidade. A sequência final para o inserto (“consensus”) foi obtida através do desenho de oligonucleotídeos de extensão (“primers walking”) para unir os dois maiores fragmentos de montagem (“contigs”) gerados pelo algoritmo de montagem (“assembler”) para que se pudesse superar uma barreira de sequenciamento contida no inserto cosmidial ocasionada por um alto conteúdo C/G.

Foram identificadas as Fases Abertas de Leitura (ORFs - *software* ORFfinder: <http://www.ncbi.nlm.nih.gov/projects/gorf/>) no inserto metagenômico, e foi feita uma triagem manual de prováveis genes com base na maior similaridade a sequências depositadas em bancos públicos de dados integrados ao NCBI (BlastP, ALTSCHEL *et al.*, 1990), seguido de correção de erros da anotação em códons de iniciação. A mesma sequência nucleotídica do “contig” foi submetida a um programa de identificação de domínios conservados: o PRODOM (SERVANT *et al.*, 2002), buscando uma confirmação da previsão funcional.

Análises de modelagem estrutural

A ORF correspondente a OMT-M (ORF20) foi submetida a modelagem estrutural por homologia (KELLEY e STERNBERG, 2009). Para tanto procedeu-se o alinhamento da sequência alvo às proteínas de estrutura terciária previamente resolvida por métodos experimentais (RMN e Cristalografia de Raio-X) disponíveis no Banco de dados de Proteínas PDB (NCBI) através do emprego do software Phyre (<<http://www.sbg.bio.ic.ac.uk/phyre/>>; KELLEY e STERNBERG, 2009). Os resultados de alinhamento gerados que retornaram uma identidade entre os moldes e a ORF inferiores a 30% foram rejeitados. Os demais foram submetidos à modelagem por homologia empregando o software Swiss-model (ARNOLD *et al.* 2006) para gerar os modelos estruturais para as ORFs, bem como para efetuar as análises de “assessing” empregando os *softwares* disponíveis na interface Swiss-model: (<http://swissmodel.expasy.org/>): Qmean, Procheck. Os modelos foram também

submetidos ao servidor Verify3D:(http://nihserver.mbi.ucla.edu/Verify_3D/).

A visualização e processamento de imagens das estruturas 3D foram realizadas empregando-se o software Pymol para o MAC-OS-X (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

4.4. RESULTADOS E DISCUSSÃO

Identificação da sequência metagenômica por homologia

A ORF metagenômica foi analisada empregando os softwares BlastP/BlastN, (ALTSCHUL *et al.*, 1990) seguido de correção de erros da anotação em códons de iniciação. A mesma sequência foi submetida a um programa de identificação de domínios conservados: o PRODOM (SERVANT *et al.*, 2002). A Tabela 3 relaciona as inferências funcionais obtidas por homologia, bem como um resumo sobre os domínios conservados encontrados.

Tabela 3 – Predição funcional por homologia e predição de domínios enzimáticos

Enzima	Homologia BlastP	SI/ %*	Domínios enzimáticos encontrados (PRODOM)
O-metil-transferase (família2) / ORF22	<i>Streptomyces</i> sp. ACTE ZP_06273109.1	85/73	1: <i>Metil-transferase Transferase O-Metil-transferase ácida de Biossíntese de lignina Caféica 3-O-Metil-transferase Família O-Metil-transferase</i> (PD001204) / 1: <i>Família de Transferase O-Metil-transferase Metil-transferase</i> (PDA7O0D6) / 1: <i>Transferase Metil-transferase O-Metil-transferase OXYF</i> (PDA2B2K5) /1: <i>Transferase O-Metil-transferase Metil-transferase</i> (PDA7H3P9).

*similaridade/Identidade

A presença dos domínios enzimáticos permite comprovar a identidade obtida pela submissão da sequência ao BlastP, indicando se tratar de uma enzima pertencente à família das alcoóis-transferases, sendo dessa forma potencialmente útil para o uso na biossíntese combinatória para a obtenção de compostos policetídicos. Sendo assim, a mesma foi submetida ao processo de construção de modelos tridimensionais por homologia descrito a seguir.

Modelagem 3D *in silico* e análises de qualidade preditiva dos modelos (“assessing”)

A modelagem estrutural da proteína alvo foi feita empregando as ferramentas disponíveis na interface do PHYRE (KELLEY e STERNBERG, 2009) para gerar um alinhamento entre esta e as sequências do banco de dados PDB e uma vez que foi detectada uma identidade superior a 30% entre a sequência metagenômica e o molde, foi realizada a predição do modelo tridimensional da álcool-transferase metagenômica através da interface Swiss-model, tendo por base a estrutura gerada por cristalografia de raio-X mais similar. A estrutura 3D gerada foi em seguida avaliada quanto à acurácia de predição, empregando as ferramentas de “assessing” Procheck, Verify3D (Tab. 4) e Qmean (Fig. 1). Como parâmetro de comparação, a estrutura do molde (2IP2) também foi submetida a essas análises.

Tabela 4 - Avaliação estereo-química molde/modelo

3D/L	Id. (%) ^a	Regiões obtidas no gráfico Ramachandran plot (Procheck) ^b				Escore 3D Profile*		
		Reg. mais favorável (%)	Reg. permitida(%)	Reg. generosamente permitida(%)	Reg. não permitida(%)	Total	Ideal	S _{ideal}
2IP2 (330)	32	93,4	6,6	0	0	161,67	150,72	1,07
OMT-M (333)		88,9	9,8	0,7	0,7	116,35	152,11	0,76

^a: resultado para o molde mais similar (foram empregados alinhamentos múltiplos), Id.: Identidade. ^b: Gráfico de Ramachandran (PROCHECK) baseado nos ângulos de torção Φ e Ψ (C-alpha e N e C-alpha e C-peptídico). *Dados calculados a partir da análise no programa Verify3 (http://nihserver.mbi.ucla.edu/Verify_3D/) Escore Total: fornecido. Escore ideal: $\text{Exp}^{-0,83+1,008 \times \ln(L)}$, e S_{ideal}, que equivale a compatibilidade da sequência com sua estrutura 3D, e é obtido por meio da divisão entre o escore total e escore Ideal, sendo que valores ideais estão acima de 0,45 S_{ideal}. L.: número de resíduos de aminoácidos (DA SILVEIRA *et al.*, 2005).

Um modelo estrutural de alta precisão (tanto modelagem experimental quanto *in silico*) deve ter em torno de 90% de seus aminoácidos localizados na região mais favorável do gráfico de **Ramachandran (PROCHECK)**. No entanto, assume-se que acima de 85% obtenha-se modelos com um bom índice de confiança (DA SILVEIRA *et al.*, 2005). Como pode ser observado na Tabela 4, o resultado obtido para a O-metil-

transferase metagenômica (OMT-M) ficou bem próximo ao valor necessário para modelos de alta acurácia: 88,9%.

A Figura 27 apresenta resultados obtidos por submissão ao **Qmean** para a resolução do modelo da OMT-T, bem como de um molde apresentando duas cadeias (C) e outro com somente uma (A), sendo que ambos foram utilizados para gerar o gráfico de distribuição atômica utilizado para a construção do modelo preditivo obtido. Pode-se perceber que as resoluções dos modelos de cadeias simples são comparáveis entre si, indicando que o modelo predito se assemelha ao obtido experimentalmente, o que também pode decorrer da necessidade de duas cadeias para uma correta determinação estereo-química funcional no modelo real. Uma observação nos resultados de S_{ideal} . (Tab.4) de todas as estruturas avaliadas permite verificar que todas assumiram valores acima de 0,45 S_{ideal} , o que permite dizer que existe uma compatibilidade entre a estrutura primária e terciária dos mesmos, não havendo portanto erros grosseiros na modelagem das estruturas preditas.

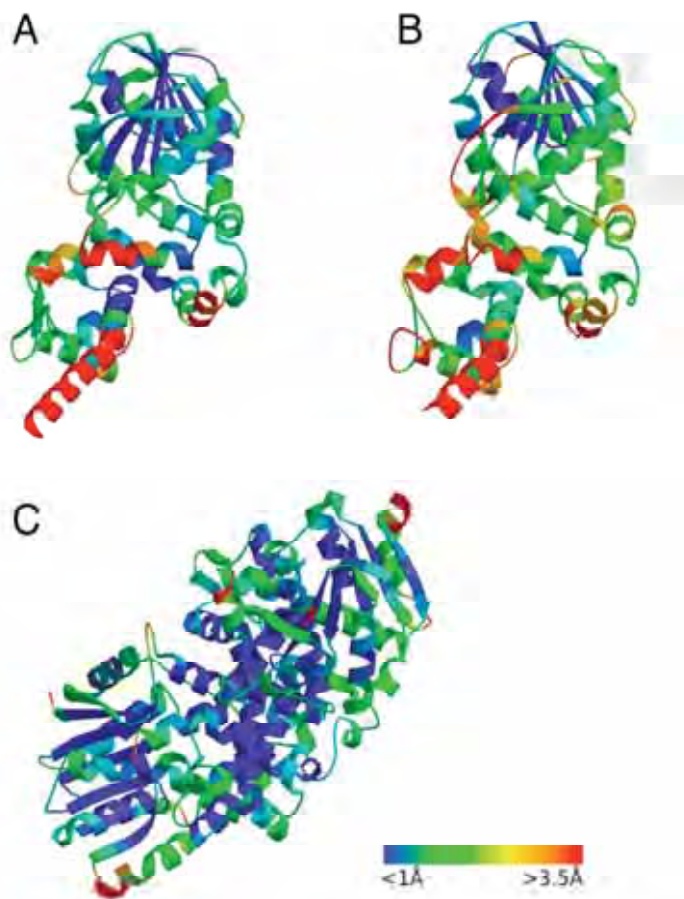


Figura 27: Predição estrutural para o molde e a OMT-M, com análise de erros de estrutura pelo Qmean. A: 3gz; B: OMT-M; C: modelo 2ip2 (cadeias A e B).

Como pode ser observado nas Figuras 27 e 28, a OMT-M apresenta-se como uma estrutura formada por cadeia única, enquanto que o molde de maior identidade (2ip2) apresenta duas cadeias. Não é possível determinar se a estrutura metagenômica é menor ou se trata de uma enzima homodimérica, em que duas cadeias polipeptídicas seriam sintetizadas para compor a estrutura completa (Estrutura quaternária). No entanto, é possível observar correspondências nas arquiteturas do modelo do molde, indicando conservação de partes da cadeia homóloga. Abaixo (Fig. 28), tem-se uma comparação entre a estrutura da OMT-M e o molde 2ip2, destacando-se uma das cadeias do molde para melhor comparação visual entre eles.

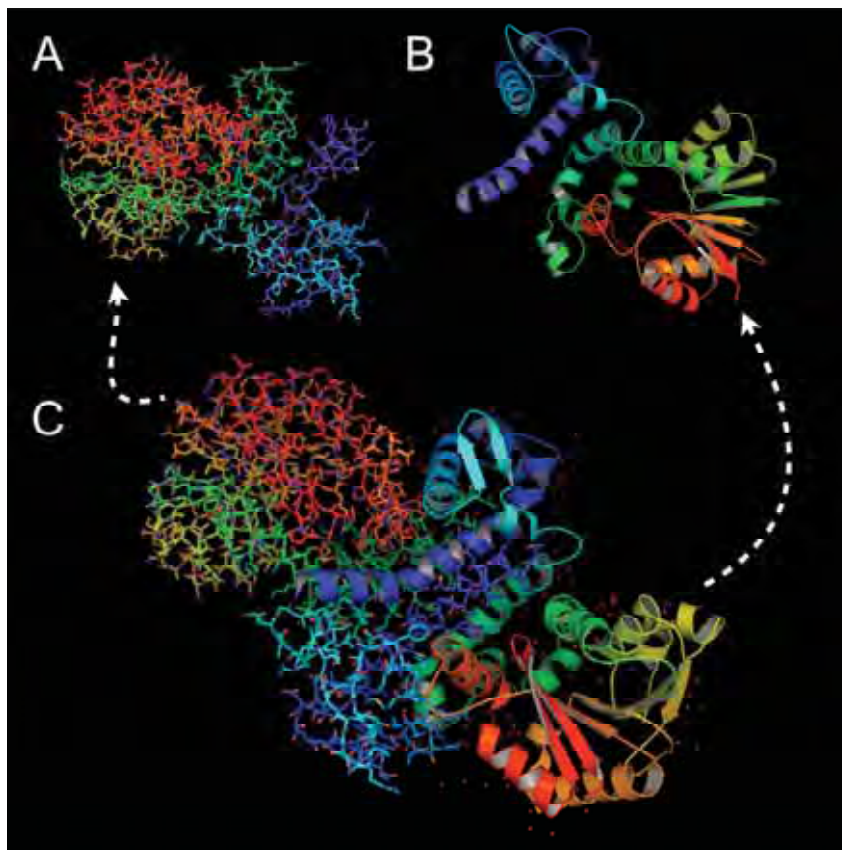


Figura 28: Conformação estrutural para o molde 2ip2 e a OMT-M, destacando as cadeias enzimáticas do molde 2ip2, com recursos Wizard do MacPymol; A: OMT-M, representação técnica da cadeia; B: OMT-M, representação no modo “Desenho”, para as estruturas secundárias. C: modelo 2ip2, cadeias A (representação técnica da cadeia) e B (representação no modo “Desenho de estruturas secundárias”).

Em ambos os modos de visualização (A e B) é possível identificar semelhanças entre o modelo gerado e a estrutura determinada por cristalografia. Contudo, o modelo da OMT-M apresenta particularidades, como pode ser visto no detalhe de folhas beta antiparalelas representadas em azul claro para o molde (C, representação no modo “Desenho” de estruturas secundárias), que estão substituídas por regiões de “loops” na OMT-M.

4.5. Conclusões

A predição funcional por ortologia e modelagem estrutural *in silico* permitiu traçar um perfil razoavelmente detalhado para uma O-metil-transferase acessada por estratégia metagenômica. A análise da confiança estatística do modelo permitiu gerar dados confiáveis, uma vez que ponderou o limite das extrapolações inferidas sobre os dados para elucidação de processos e funções metabólicas.

A enzima encontrada possui um grande potencial para a aplicação na biossíntese combinatória para a produção de antibióticos, dentre outros compostos de interesse biotecnológico.

CAPÍTULO 5 – Implicações

Os estudos envolvendo Metagenômica permitem o acesso praticamente irrestrito a um imenso reservatório gênico inexplorado.

Diante desta imensa riqueza, o grande desafio não é só encontrar novos genes: é preciso entender qual o papel dos mesmos no ambiente em que foi isolado (fator ecológico), lançando luz sobre quais aplicações biotecnológicas ele possui.

Outro desafio imposto pela metagenômica, visto para muitos pesquisadores como o “gargalo” para a aplicação dessa ferramenta, é a adequação do correto hospedeiro heterólogo.

No caso específico de produção de antibióticos por vias de PKS II, além deste impasse, é necessário identificar qual o caminho mais promissor para se chegar a um composto de interesse premente, acertando na escolha das enzimas que serão recombinadas, prevendo sua capacidade de atuação conjunta.

A metodologia empregue neste estudo explorou o conceito de homologia para inferir as funcionalidades catalíticas para os produtos de expressão dos genes descobertos por metagenômica. Em conjunto, as estratégias utilizadas permitiram recuperar uma via não previamente sequenciada, e estudar as hipóteses de homologia

de uma forma detalhada, devido sobretudo à identificação de domínios conservados enzimáticos, e ao cruzamento destas informações com dados previamente publicados.

O uso da filogenia permitiu visualizar como o “cluster” metagenômico se relaciona com outras vias diversificadas, através da comparação entre regiões gênicas conservadas de KSAs. Uma compilação de mapas enzimáticos de complexos de PKS II da literatura, envolvidos tanto com a PKS de produção de pigmentos, quanto com as da produção de antibióticos, permitiu um interessante parâmetro comparativo, para avaliar o “cluster” estudado.

O uso da modelagem estrutural por homologia, por sua vez, permitiu um “arremate” das inferências funcionais obtidas, permitindo simular o dobramento das proteínas relacionadas aos genes encontrados, além da comparação visual dos modelos 3D metagenômicos obtidos, com estruturas 3D previamente elucidadas por Cristalografia de Raio-X ou Ressonância Magnética Nuclear.

Os resultados obtidos irão subsidiar estudos posteriores em expressão gênica heteróloga, e serão úteis para a determinação do delineamento experimental mais promissor para se chegar à obtenção dos produtos de expressão enzimática para caracterização da cinética enzimológica da via.

Os modelos 3D gerados serão empregues em estudos preditivos de interação enzimática, sobretudo em estratégias de “docking” proteína-proteína e proteína-ligante, buscando determinar quais os substratos enzimáticos, e estudar a compatibilidade das subunidades catalíticas com enzimas de outras vias. Uma vez validadas as estratégias *in silico* por meio de experimentos *in vitro* e *in vivo* (hospedeiros heterólogos), a metodologia padronizada será utilizada para gerar um banco de dados interno do LBMP para as enzimas de PKS metagenômicas, visando facilitar o reconhecimento de unidades enzimáticas promissoras para o desenho racional de novos fármacos.

Outras enzimas da via possuem potencial de aplicação biotecnológica em outras áreas (celulases, lacases, peptidases, oxidases, dentre outras) e serão estudadas em paralelo empregando-se as abordagens descritas.

6. REFERÊNCIAS

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, London, v.10, 215-403, 1990.

AMANN, R.I.L.; LUDWIG, W.; SCHLEIFER, K.H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. **Microbiological Reviews**, Washington, v. 59, 143-169, 1995.

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

BAERSON, S.R. & RIMANDO, A.M. A Plethora of Polyketides: Structures, Biological Activities, and Enzymes In: **American Chemical Society**, U.S. government work, Washington, 2007.

BALTZ, H. Molecular engineering approaches to peptide, polyketide and other antibiotics. **Nature Biotechnology**, London, v. 24, 1533-1540, 2006.

BENKERT, P.; KUENZLI, M. & SCHWEDE, T. QMEAN Server for Protein Model Quality Estimation. **Nucleic Acids Research**, Oxford, v.37, W510-4, mai. 2009.

BETTIGA, G.H., GORWA-GRAUSLUND & HAHN-HAGERDAL. Metabolic Engineering in Yeast. *In*: HandBook of metabolic pathways, v. 22, 680 p., 2010.

BODE, H.B. & MÜLLER, R. The impact of bacterial genomics on natural product research, **Angewandte chemie**, Weinheim, v.44, 6828-6846, 2005.

BOWIE, J.U., LUTHY, R. & EISENBERG, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. **Science**, Washington, v.253, 164-253, julh. 1992.

BRACHMANN, A. O.; JOYCE, S.A.; JENKE-KODAMA, H.; SCHWÄR, G.; CLARKE, D.J.; BODE, H.B. A Type II Polyketide Synthase is Responsible for Anthraquinone Biosynthesis in *Photorhabdus luminescens*. **ChemBioChem**, Weinheim, v.8, 1721-1728, 2007.

BULL, A.T. ; WARD, A.C.; GOODFELLOW, M. Search and discovery strategies for biotechnology: the paradigm shift. **Microbiology and Molecular Biology Reviews**, New York, v.64, p.573-606, 2000.

BUSTI, E.; MONCIARDINI, P.; CAVALETTI, L.; BAMONTE, R.; LAZZARINI, A.; MARGHERITA SOSIO, M.; DONADIO, S. Antibiotic - producing ability by representatives of a newly discovered lineage of actinomycetes. **Microbiology**, Great Britain, v.152, 675-683, 2006.

BUTLER, M.J.; BRUHEIM, P.; JOVETIC, S.; MARINELLI, F.; POSTMA, P.W.; BIBB, M.J. Engineering of Primary Carbon Metabolism for Improved Antibiotic Production in *Streptomyces lividans*. **Applied and Environmental Microbiology**, Washington, v. 68, 4731-4739, 2008.

CASTALDO, G.; ZUCKO, J.; HEIDELBERGER, S.; VUJAKLIJA, D.; HRANUELI, D.; CULLUM, J.; WATTANA-AMORN, P.; CRUMP, M.P.; CROSBY, J.; LONG, P.F. Proposed Arrangement of Proteins Forming a Bacterial Type II Polyketide Synthase. **Chemistry & Biology**, Cambridge, v.15, 1156–1165, nov. 2008.

CHALLIS, G.L. Mining microbial genomes for new natural products and biosynthetic pathways. **Microbiology**, Great Britain, v.154, 1555-1569, 2008.

COMMITTEE ON METAGENOMICS, Board of Life Sciences. The New Science of Metagenomics Revealing the Secrets of Our Microbial Planet. **The National Academies Press**, 2007.

COURTOIS, S.; CAPPELLANO, C.M.; BALL, M.; FRANCOU, F-X.; NORMAND, P.; HELYNCK, G.; MARTINEZ, A.; KOLVEK, S.J.; HOPKE, J.; OSBURNE, M.S.; AUGUST, P.R.; NALIN, R.; GUÉRINEAU, M.; JEANNIN, P.; SIMONET, P.; PERNODET, J-L.

Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. **Applied and Environmental Microbiology**, Washington, v.69, 49-55, 2003.

COWAN, D.; MEYER, Q.; STAFFORD, W.; MUYANGA, S.; CAMERON, R.; WITTEW, P. Metagenomic gene discovery: past, present and future. **TRENDS in Biotechnology**, Amsterdam, v.23, 321-329, 2005.

D'SOUZA-TICLO, D. T. The lignin-Degrading enzyme, Laccase from Marine Fungi; Biochemical and Molecular Approaches. Tese de Doutorado apresentada ao National Institute of Oceanography. (2008)

DA SILVEIRA, N.J.F.; UCHÔA, H.B.; PEREIRA, J.H.; CANDURI, F.; BASSO, L.A.; PALMA, M.S.; SANTOS, D.S.; AZEVEDO Jr., W.F. Molecular models of protein targets from *Mycobacterium tuberculosis*. **Journal of Molecular Modeling**, Berlin, v.11, 160-166, 2005.

DAVIES, J. Are antibiotics naturally antibiotics? **Journal of Industrial Microbiology and Biotechnology**, Hampshire, v.33, 496-499, 2006.

DICK, G.J.; TORPEY, J.W.; BEVERIDGE, T.J.; TEBO, B.M. Direct Identification of a Bacterial Manganese (II) Oxidase, the Multicopper Oxidase MnxG, from Spores of Several Different Marine Bacillus Species. **Applied and Environmental Microbiology**, Washington, v.74, 1527-1534, 2008.

EMBRAPA, DF. O programa Blast: Guia prático de utilização. Brasília, Embrapa Recursos Genéticos e Biotecnologia, 2007.

EWING, B. & GREEN, P. Base-calling of automated sequencer traces using Phred II Error probabilities. **Genome Research**, Cold Spring Harbor, v.8, 186-194, 1998.

EWING, B.; HILIER, L.D; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. **Genome Research**, Cold Spring Harbor, v.8, 175-185, 1998.

FELSENSTEIN, J. Phylip (Phylogeny Inference Package) version 3.6. **Distributed by the author**. Department of Genome Sciences, University of Washington, Seattle, 2005.

FINDLOW, S.C.; WINSOR, C.; SIMPSON, T.J.; CROSBY, J.; CRUMP, M.P. Solution Structure and Dynamics of Oxytetracycline Polyketide Synthase Acyl Carrier Protein from *Streptomyces rimosus*. **Biochemistry**, Washington, v.42, 8423-8433, 2003.

FINN, R.D.; MISTRY, J.; TATE, J.; COGGILL, P.; ANDREAS HEGER, A.; JOANNE E. POLLINGTON, J.E.; GAVIN, O.L.; GUNASEKARAN, P.; CERIC, G.; FORSLUND, K.; HOLM, K.L.; SONNHAMMER, E.L.L.; EDDY, S.R.; BATEMAN, A. The Pfam protein families database. **Nucleic Acids Research**, Oxford, v. 38, D211-222, nov. 2010.

FISCHBACH, M. A., WALSH, C. T. & CLARDY, J. The evolution of gene collectives: How natural selection drives chemical innovation. **PNAS**, Washington, v.105, 4601-4608, 2008.

FISCHBACH, M. A. & WALSH, C. T. Antibiotics for Emerging Pathogens. **Science**, Washington, v.325, 1089-1093, 2009.

GENILLOUD, O.; GONZÁLEZ, I.; SALAZAR, O.; MARTÍN, J.; TORMO, J. R.; VICENTE, F. Current approaches to exploit actinomycetes as a source of novel natural products. **Journal of Industrial Microbiology and Biotechnology**, Hampshire, Doi: 10.1007/s10295-010-0882-7, 2010.

GOMES, E.S.; NAVARRETE, A.A.; LEMOS, E.G.M.; TSAI, S.M.; MOREIRA, F.M.S. A nova ciência da metagenômica: revelando os segredos do planeta microbiano. **Boletim informativo da SBCS**, Viçosa, v. 34, 20-22, jan-abr. 2010.

GOMES, E. S. Utilização da metagenômica como ferramenta para prospecção de novos genes do sistema PKS tipo II. Trabalho apresentado à Faculdade de Ciências Agrárias e Veterinárias – UNESP, Campus de Jaboticabal, para obtenção do título de Bacharel em Ciências Biológicas, 2008.

GORDON, D., ABAJIAN, C. & GREEN, P. Consed: A Graphical Tool For Sequence Finishing. **Genome Research**, Cold Spring Harbor, v.8, 195-202, 1998.

GRAZIANI, S., VASNIER, C. & DABOUSSI, M. Novel Polyketide Synthase from *Nectria haematococca*. **Applied And Environmental Microbiology**, Washington, v.70, 2984–2988, 2004.

GROSS, F.; LUNIAK, N.; PERLOVA, O.; GAITATZIS, N.; HOLGER JENKE-KODAMA, H.; GERTH, K.; GOTTSCHALK, D.; DITTMANN, E.; ROLF MÜLLER, R. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. **Archives of Microbiology**, Berlin, v. 185, p. 28-38, 2006.

GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. Codon bias and heterologous protein expression. **TRENDS in Biotechnology**, Amsterdam, v.22, 2004.

HAMAD, B. The antibiotics market. **Nature Drug Discovery**, London, v.9, 675-676, 2010.

HE, W.; LEI, J.; LIU, Y.; WANG, Y. The LuxR family members GdmRI and GdmRII are positive regulators of geldanamycin biosynthesis in *Streptomyces hygroscopicus* 17997. **Archives of Microbiology**, Berlin, v.510, 189-501, 2008.

HOLGER, J.H. & DITTMANN, E. Evolution of metabolic diversity: Insights from microbial polyketide synthases. **Phytochemistry**, New York, v.70, 1858–1866, 2009.

HOPWOOD, D. A.; MALPARTIDA, F.; KIESER, H. M.; IKEDA, H.; DUNCAN, J.; FUJII, I.; RUDD, B. A. M.; FLOSS, H. G.; OMURA, S. Production of 'hybrid' antibiotics by genetic engineering. **Nature**, London, v.314, 642-644, 1985.

HOPWOOD, D. A. Genetic Contributions to Understanding Polyketide Synthases. **Chemical Reviews**, Washington, v.97, 2465-2497, 1997.

HSIAO, N. & KIRBY, R. Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a

Streptomyces coelicolor microarray system. **Antonie van Leeuwenhoek**, Amsterdam, v. 93, 1–25, 2008.

HUANG, J.; SHI, J.; MOLLE, V.; SOHLBERG, B.; WEAVER, D.; BIBB, M.J.; KAROONUTHAISIRI, N.; LIH, C-J.; KAO, C.M.; BUTTNER, M.J.; COHEN, S.N. Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. **Molecular Microbiology**, Salem, v. 58, 1276-1287, 2005.

HUTCHINSON, C.R. & FUJII, I. Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. **Annual Review of Microbiology**, Palo Alto, v.49, 201-238, 1995.

JENKE-KODAMA, H. & DITTMANN, E. Evolution of metabolic diversity: Insights from microbial polyketide synthases. **Phytochemistry**, New York, v.70, 1858-1866, 2009.

JOURNAL OF ANTIBIOTICS. Special issue of The Journal of Antibiotics dedicated to the late Professor C Richard Hutchinson, **Journal of Antibiotics**, Tokyo, v. 64, 3-5, 2011.

KAITIN, K. Deconstructing the Drug Development Process: The New Face of Innovation. **Clinical Pharmacology & Therapeutics**, Saint Louis, v. 87, 356-361, 2010.

KEATINGE-CLAY, A.T.; SHELAT, A.A.; SAVAGE, D.F.; TSAI, S.C., MIERCKE, L.J.; O'CONNELL, J.D. III; KHOSLA, C.; STROUD, R.M. Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA:ACP trans-acylase. **Structure**, Philadelphia, v.11, 147-154, 2003.

KELLEY, L.A. & STERNBERG, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. **Nature Protocols**, London, v.4, 363-371, fev. 2009.

KNELLER, R. The importance of new companies for drug discovery: origins of a decade of new drugs. **Drug Discovery**, London, v. 9, 867-882, 2010.

LAM, S.; GRACE, A.H.; MATTEI, J. M.; STEPHEN, W.M.; FORENZA, S. Hymastatin, a New Antitumor Antibiotic from *Streptomyces hygroscopicus*. **The Journal of antibiotics**, Tokyo, v.43, 956-960, 1990.

LARKIN, M.A.; BLACKSHIELDS, G.; BROWN, N.P.; CHENNA, R.; MCGETTIGAN, P.A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I.M.; WILM, A.; LOPEZ, R.; THOMPSON, J.D.; GIBSON, T.J.; HIGGINS, D.G. Clustal W and Clustal X version 2.0. **Bioinformatics**, Oxford, v.23, 2947-2948, 2007.

LASKOWSKI, R.A.; MACARTHUR, M.W.; MOSS, D.S.; THORNTON, J.M. PROCHECK: A program to check the stereochemical quality of protein structures, **Journal of Applied Crystallography**, Copenhagen, v.26, 283-291, 1993.

LILES, M.R.; WILLIAMSON, L.L.; RODBUMRER, J.; TORSVIK, V.; GOODMAN, R.M.; HANDELSMAN, J. Recovery, Purification, and Cloning of High Molecular-Weight DNA from Soil Microorganisms. **Applied And Environmental Microbiology**, Washington, v.74, 3302–3305, 2008.

LOVELL, S.C.; DAVIS, I.W.; ARENDALL, W.B. III; DE BAKKER, P.I.; WORD, J.M.; PRISANT, M.G.; RICHARDSON, J.S.; RICHARDSON, D.C. Structure Validation by C α Geometry: ϕ , ψ and C β Deviation. **Proteins: Structure, Function, and Genetics**, Hoboken, v.50, 437-450, 2003.

LÜTHY, R., BOWIE, J.U. & EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, London, v. 356, 83-85, 1992.

MARMUR, J. Procedure for the Isolation of Deoxyribonucleic Acid from Microorganisms. **Journal of Molecular Biology**, London, v.3, 208-218, 1961.

MARTI-RENOM, M.A.; MADHUSUDHAN, M.S.; FISER, A.; ROST, B.; SALI, A. Reliability of assessment of protein structure prediction methods. **Structure**, Philadelphia, v.10, 435-440, 2002.

MARTIN, J.F. & DEMAINE, A.L. Control of Antibiotic Biosynthesis, **Microbiological Reviews**, Washington, v. 44, 230-251, 1980.

MASSINI, K.C. Bioprospecção de Genes Biossintéticos de Policetídeos em DNA Metagenômico de Solo de Mata Atlântica. Tese apresentada ao Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências (Microbiologia). São Paulo, 2009, 114 pp.

MÉNDEZ, C. & SALAS, J.A. ABC transporters in antibiotic-producing actinomycetes. **FEMS Microbiology Letters**, Amsterdam, v.158, 1-8,1998.

METSÄ-KETELÄ, M. ; SALO, V.; HALO, L.; HAUTALA, A.; HAKALA, J.; MÄNTSÄLÄ, P.; YLIHONKO, K. An efficient approach for screening minimal PKS genes from *Streptomyces*. **FEMS Microbiology Letters**, Amsterdam, v.180, 1-6, 1999.

MOTCHE, M.; DEHESH, K.; EDWARDS, P.; LINDQVIST, Y. The Crystal Structure of β -Ketoacyl-Acyl Carrier Protein Synthase II from *Synechocystis sp.* at 1.54 Å Resolution and Its Relationship to Other Condensing Enzymes. **Journal of Molecular Biology**, London, v. 305, p. 491-503, 2001.

MULDER, N.J.; Apweiler, R; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bradley, P.; Bork, P.; Bucher, P.; Cerutti, L.; Copley, R.; Courcelle, E.; Das, U.; Durbin, R.; Fleischmann, W.; Gough, J.; Haft, D.; Harte, N.; Hulo, N.; Kahn, D.; Kanapin, A.; Krestyaninova, M.; Lonsdale, D.; Lopez, R.; Letunic, I.; Madera, M.; Maslen, J.; McDowall, J.; Mitchell, A.; Nikolskaya, A.N.; Orchard, S.; Pagni, M.; Ponting, C.P.; Quevillon, E.; Selengut, J.; Sigrist, C.J.; Silventoinen, V.; Studholme, D.J.; Vaughan, R.; Wu, C.H. InterPro, progress and status in 2005. **Nucleic Acids Research**, Oxford, v.33, D201-205, 2005.

OLANO, C. Hutchinson's legacy: keeping on polyketide biosynthesis. **The Journal of Antibiotics**, Tokyo, v. 64, 51–57, 2011.

OMURA, S.; IKEDA, H.; ISHIKAWA, J.; HANAMOTO, A. TAKAHASHI, C.; SHINOSE, M; TAKAHASHI, Y.; HORIKAWA, H.; NAKAZAWA, H.; OSONOE, T.; KIKUCHI, H.; SHIBA,

T.; SAKAKI, Y.; HATTORI, M. Genome sequence of an industrial Microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. **Proceedings of the National Academy of Sciences** (Usa), Washington, v.98, 12215-12220, 2001.

PEIRÚ, S.; RODRÍGUEZ, E.; MENZELLA, H.G.; CARNEY, J.R.; GRAMAJO, H. Metabolically engineered *Escherichia coli* for efficient production of glycosylated natural products. **Microbial Biotechnology**, v. 1, 476–486, 2008.

PERRIÈRE, G & GOUY, M. WWW-Query: An on-line retrieval system for biological sequence banks. **Biochimie**, Paris, v.78, 364-369, 1996.

PEREIRA, M. R. Prospecção de genes codificadores de enzimas lipolíticas em biblioteca metagenômica de consórcio microbiano degradador de óleo diesel. 2011. Dissertação (Mestrado em Biotecnologia) - Universidade de São Paulo, São Paulo.

PIDOT, S.J.; HONG, H.; SEEMANN, T.; PORTER, J.L.; YIP, M.J.; MEN, A.; JOHNSON, M.; WILSON, P.; DAVIES, J.K.; LEADLAY, P.F.; STINEAR, T.P. Deciphering the genetic basis for polyketide variation among mycobacteria producing mycolactones. **BMC Genomics**, London, v.9, p. 462, 2008.

PROSDOCIMI, F.; CERQUEIRA, G.C.; BINNECK, E.; SILVA, A.F.; REIS, A.N.; JUNQUEIRA, A.C.M.; SANTOS, A.C.F.; NHANI-JÚNIOR, A.; WUST, C.I.; CAMARGO-FILHO, F.; KESSEDJIAN, J.L.; PETRETSKI, J.H.; CAMARGO, L.P.; FERREIRA, R.G.M.; LIMA, R.P.; PEREIRA, R.M.; JARDIM, S.; SAMPAIO, V.S.; FOLGUERAS-FLATSCHART, A.V. BIOINFORMÁTICA: Manual do usuário. Biotecnologia Ciência & Desenvolvimento, Data de acesso: 10/02/2009, Disponível em: <http://www.icb.ufmg.br/prodabi/prodabi5/homepages/liza/bioinf.pdf>.

QUEVILLON, E.; SILVENTOINEN, V.; PILLAI, S.; HARTE, N.; MULDER, N.; APWEILER, R.; LOPEZ, R. InterProScan: protein domains identifier. **Nucleic Acids Research**, Oxford, v.33, W116-w120, 2005.

RIX, U.; FISCHER, C.; REMSING, L.L.; ROHR, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. **Natural Product Reports**, London, v.19, 542-580, 2002.

SAMBROOK, J. & RUSSEL, D.W. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001.

SCHMEISSER, C., STEELE, H. & STREIT, W.R. Metagenomics, biotechnology with non-culturable microbes. **Applied Microbiology and Biotechnology**, Berlin, v.75, 955-962, 2007.

SCHUCH, V. Prospecção de genes envolvidos na biossíntese de antibióticos através da abordagem metagenômica. Dissertação de Mestrado em Microbiologia. Faculdade de Ciências Agrárias e Veterinárias, UNESP. Jaboticabal, 2007.

SCHUCH, V., GOMES, E.S. & LEMOS, E.G.M. Discovery of genes related to antibiotic biosynthesis into a metagenomic library. **Elsevier-New Biotechnology**, v. 25, 2009.

SERVANT, F.; BRU, C.; CARRÈRE, S.; COURCELLE, E.; GOUZY, J.; PEYRUC, D.; KAHN, D. ProDom: Automated clustering of homologous domains. **Briefings in Bioinformatics**, London, v.3, 246-251, set. 2002.

SHEN, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms . **Current Opinion in Chemical Biology**, v.7, 285–295, 2003.

SHEN, Y.; YOON, P.; YU, T.W.; FLOSS, H.G.; HOPWOOD, D.; MOORE, B.S. Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. **Biochemistry**, Washington, v.96, 3622–3627, 1999.

SILVEIRA, E. L.; Pereira, R.M.; Scaquitto, D.C.; Pedrinho, E.A.N.; Val-Moraes, S.P.; Wickert, E.; Carareto-Alves, L.M.; Lemos, E.G.M. Bacterial diversity of soil under eucalyptus assessed by 16S rDNA sequencing analysis. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, 1507-1516, 2006.

SJÖLANDER, K. Phylogenomic inference of protein molecular function: advances and challenges. **Bioinformatics**, Oxford, v.20, 170-179, 2004.

SMITH, T. F. & WATERMAN, M. S. Identification of common molecular subsequences. **Journal of Molecular Biology**, London, v. 147, 195-197, 1981.

SOSIO, M. & DONADIO, S. Understanding and manipulating glycopeptides pathways: the example of the dalbavancin precursor A40926. **J. Ind. Microbiol Biotechnol**, v.33, 569-576, 2006.

TOVCHIGRECHKO, A. & VAKSER, I. A. Gramm-X Public Web Server For Protein-Protein Docking. **Nucleic Acids Research**, Oxford, v.34, W310-W314, Web Server issue doi:10.1093/nar/gkl206, 2006.

TSAI, S-C. S. & AMES, B.D. Structural Enzymology of Polyketide Synthases. In: Weissman, K. J., Introduction to Polyketide Biosynthesis. **Methods Enzymol**, v.459, 3-16, 2009.

VIDOTTI, C. C. F.; DE CASTRO, L. L. C. & CALIL, S. S. New drugs in Brazil: Do they meet Brazilian public health needs? **Rev Panam Salud Publica**, v.24, 36-45, 2008.

WADT, M. Análise Econômica de Novos Fármacos Licenciados no Brasil entre 1998 e 2001. Dissertação apresentada à Faculdade de Ciências Farmacêuticas da Universidade de São Paulo, para obtenção do Título de Mestre em Farmacologia. São Paulo, 2003, 236 pp.

Wang, G.Y.; Graziani, E.; Waters, B.; Pan, W.; Li, X.; McDermott, J.; Meurer, G.; Saxena, G.; Andersen, R.J.; Davies, J. Novel natural products from soil DNA libraries in a *Streptomyces* host. **Organic Letters**, Washington, v.2, 2401-2404, 2000.

WANG, X-J.; WANG, X-J; YAN, Y-J.; ZHANG, B.; AN, J.; WANG, J-J.; TIAN, J.; JIANG, L.; CHEN, Y-H; HUANG, S-X; YIN, M.; ZHANG, J.; GAO, A-L.; LIU, C-X; ZHU, Z-X.X; XIANG, W-S. Genome Announcement: Genome Sequence of the Milbemycins-

producing Bacterium *Streptomyces bingchenggensis*. **Journal of Bacteriology**, Washington, v.192, 4526-4527, 2010.

WAWRIK, B.; KERKHOF, L.; ZYLSTRA, G.J.; KUKOR, J.J. Identification of Unique Type II Polyketide Synthase Genes in Soil. **Applied And Environmental Microbiology**, Washington, v.71, 2232-2238, 2005.

WEISSMAN, K. J. (coord.), Introduction to Polyketide Biosynthesis. **Methods Enzymol**, v.459, 2009.

WESTBROOK, J. ; FENG, Z.; CHEN, L.; YANG, H.; BERMAN, H.M. The Protein Data Bank and structural genomics. **Nucleic Acids Research**, Oxford, v.31, 489-491,2003.

WIEDERSTEIN, M. & SIPPL, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, Oxford, v.35, W407-W410, abr. 2007.

WWW.SWISS MODEL. Disponível em: < <http://swissmodel.expasy.org/course/course-index.htm> > Data de acesso: 20 de Dezembro de 2010.

YADAV, G., GOKHALE, R.S. & MOHANTY, D. SEARCHPKS: a program for detection and analysis of polyketide synthase domains. **Nucleic Acids Research**, Oxford, v.31, 3654-3658, 2003.

YU, T-W.; Shen, Y.; McDaniel, R.; Floss, H.G.; Khosla, C.; Hopwood, D.A.; Moore, B.S. Engineered Biosynthesis of Novel Polyketides from *Streptomyces* Spore Pigment Polyketide Synthases. **American Chemical Society**, Easton, v.120, 7749-7759,1998.

ZAEHNER, H. & FIELDLER, H.P. The need for new antibiotics: possible ways forward, p 67-84. In Russel, N.J. (ed). Fifty years of antimicrobials: past perspectives and future trends. **Cambridge University Press**, Cambridge, England, 1995.

ZALETA-RIVERA, K.; CHARKOUDIAN, L.K.; RIDLEY, C.P.; KHOSLA, C. Cloning, Sequencing, Heterologous Expression, and Mechanistic Analysis of A-74528 Biosynthesis. **American Chemical Society**, Easton, v.132, 9122-9128, 2010.

ZAWADA, R. J. X. & KHOSLA, C. Domain Analysis of the Molecular Recognition Features of Aromatic Polyketide Synthase Subunits. **The Journal of Biological Chemistry**, 272: 16184 –16188, 1997.

ZHANG, W. & TANG. *In Vitro* Analysis of Type II Polyketide Synthase *In*: WEISSMAN, K. J., Introduction to Polyketide Biosynthesis. **Methods Enzymology**, New York, v.459, 3-16, 2009.

ZHANG, W.; FERREIRA, J.P. & TANG, Y. Regulation of Secondary Metabolism in Bacteria. *In*: SMOLKE, C. D. (Edit.) The metabolic pathway engineering handbook : fundamentals, 1 ed., cap. 13, pg 310, 2010.

APÊNDICE

Tabela A1- Resultados obtidos por homologia utilizando o software e ProDom

ORF	Dominios (ProDom)
ORF1*	1: <i>Hidrolase Secretada</i> (PD844476).
ORF2	1: <i>Regulador de Transcrição ligante de DNA, Família LuxR Lp_1789</i> Contendo Regulador (PD337619).
ORF3	2: <i>Acil-transferase/ Transferase TCP24 Sinal precursor DBV8</i> (PD846045; PD846045).
ORF4	8: Um domínio para <i>DNA Polimerase Dna-Directed III /Transferase Hidrolase Exonuclease Epsilon Nuclease de Replicação</i> (PDA4Z479) e sete relacionados a <i>Família de Reguladores e Sítios Repressores LACI</i> (PDA7H8I4; PD000591 ; PD539176; PD275683; PD759496; PD166591 ; PD023303).
ORF5	8: Três relacionados a <i>Lipoproteínas Extracelulares Ligantes de Solutos Precursores de Sinal</i> (PDA39930; PD329227; PDA5F002); cinco relacionadas a <i>Lipoproteína Ligante de Solutos/Açúcar Específicas do Sistema ABC Transportador</i> (PD115161; PDA1X875; PD312731; PD009806; PD323955).
ORF6	6: Quatro relacionadas a <i>Regiões do Componente Inter Membrana do Sistema de Transporte ABC/ Trans-membrana-Permease</i> (PD601592; PDA112S3; PD015756; PD955326; um de <i>Trans-membrana CEBF</i> (PD758770); um de <i>ABC Sistema Transportador de SN-Glicerol-3-Phosphate/Trans-membrana Permease Plasmídeo</i> (PDA30138).
ORF7	6: Todos relacionados ao <i>Sistemas Componentes Inter-membrana Trans-membrana Dependente de Proteína Ligante</i> (PDA5C146; PD079695; PD955189; PD015756; PD865633; PDA4N621;
ORF8	1: Relacionado à <i>Hidrolase Beta-Glucosidase/Família 6-Phospho-Beta- Glicosidase /Glucosidase Hidrolase sinal precursor Glicosil-Glicosídeo</i> (PD235643).
ORF9	6: uma <i>Açúcar-hidrolase Secretada</i> (PDA9A624) / três de <i>Hidrolase quitinase transmembrana/Sinal Precursor Glicosidase</i> " (PD563565 ; PD008535 ; PD115970). Uma <i>Celulose-ligante Família tipo II bacteriana</i> (PDA9F4Z5) / uma <i>Sinal de degradação glicosidase precursora de hidrolase/ Metabolismo do Carboidrato de Celulose: Celulase quitinase</i> (PD001333).
ORF10	4: dois de <i>Sinal precursor</i> (PDA962R5; PDA962R5) / um de <i>Família de Sortase A Peptidase C60/ Sinal Precursor B similar à Sortase Associada a fimbrias</i> (PD068544)/ um <i>Precursora de Peptidase Lipoproteína A secretada/ Sortase C60 B sinal de membrana</i> (PD329252).
ORF11	3: dois de <i>Fator de Coagulação Ligante/Sinal V Repetitivo Cobre-ferro metal Precursor Oxidoreductase</i> (PD683193 ; PD186148)/um <i>Oxireductase ligante de metal cobre/Oxidase Tipo Multicopper EPOA Periplasmica Repetitiva</i> (PD821758).
ORF12	12 : dois <i>Mono-oxigenase Oxidoreductase Flavoproteína Ubiquinona Hidroxilase de Biosíntese de FAD/ de 2-Octaprenil-6-Metoxifenol e Ubih/Ubif/Visc/Coq6</i> (PD272302 ; PD000900); dois dominios similares ao anterior ubiquinona : um <i>sinal precursor Geranilgeranil</i> (PD002529) e um representante gérico (PDA9E5Q8). três : <i>Hidroxilase Oxidoreductase Mono-oxigenase Policetídeo Flavoproteína Fad de policetídeos</i> , sendo dois similares ao plasmídeo 24-Dicloro-fenol e um relacionado ao <i>Phea/Tfdb</i> (Pd005824 ; Pd455221 ; PD190945). Dois <i>Policetídeos Oxidoreductase 1.14.13.- Flavoproteína Fad Orf WHIE PKSA VIII</i> sendo um <i>hidroxilase</i> (PDA1X1J0 ; PD471436). Um <i>Oxigenase Mono-oxigenase Oxidoreductase Fad-ligante Oxigenase-Redutase de Rifampina similar à Oxigenase dependente de FAD ciclina Hidroxilase</i> (Pd338447) e duas regiões genericas relacionadas: a <i>mono-oxigenases e Fad-ligante-3</i> , (PD120511; PDA8H8D8).

Tabela A2- Resultados obtidos por homologia utilizando os software ProDom (continuação).

ORF	Dominios (ProDom)
ORF13	4: um Mono-oxigenase de Biosíntese de Antibiótico Anthrone Oxigenase WhiE Transmembrana GLL3831 AKNX I (PD291322)/ um Mono-oxigenase de Biosíntese de Antibiótico Orf Rubt (PD581598)/ dois WhiE I Whiesai Orf Locus SCHA CURD homólogo à Síntase de antibiótico (PD330905; PD024843).
ORF14	4: um Tetracenomicina WHIESAII Homóloga à Orf Schb de Biosíntese de antibiótico Policetídeo (PDA0F0P8) / um Regulador Transcricional Ligante de DNA da Família de Motif Helix-Turn-Helix/ Motif de reguladores:Cupin Cupin Xre” (PD848665) / um Tcmj Sintase de Biosíntese WHIE II homóloga ao Cupin Elmj do antibióticoTetracenomicina (PDA0I316) / um “Homólogo WHIESAII WHIE II” (PDA2Y7I5).
ORF15	10: Três domínios “Síntase Transferase 3-Oxoacil-Acíl-Carrier-Protein II Beta-Ketoacil Aciltransferase Precursor Beta-Ketoacil-Acp I (PD483857; PDA1F6K2; PDA1F6P1). Três domínios relacionados a Transferase Síntase Acyl-transferase Policetídeo Beta-Ketoacil 3-Oxoacil-Acíl-Carrier-Protein I, sendo dois relacionados a antibiótico tipo II e um policetídeo tiolase (PD000145; PD000224; PD580608). Um domínio similar para Beta-cetoacil Ketosíntase Alpha de antibióticos (PDA1G2U9). Um domínio relacionado as Síntases 3-Oxoacil-Acp tipo I (PD571518) e dois relativos ao domínio: Transferase Síntase Nonk Beta-Ketoacyl (PD297019; PD297019)
ORF16	14: Sete domínios relacionados a Beta-Ketoacil Acil-transferase ORF 2.3.1.de antibióticos, sendo uma relativa a nodulação II transmembrana; um de KS de condensação e uma CLF (PD946878; PDA1V4C3; PDA833V5; PDA1U6B8; PDA0J8R7; PDA4C9I0; PDA4C9I0). Um domínio relacionado a Síntase transferase Beta-Ketoacil Type II 3-Oxoacil-Acíl-Carrier-Proteína (PD300844); dois domínios Síntase Transferase 3-Oxoacil-Acíl-Carrier-Proteína II Beta-Ketoacil Aciltransferase I Beta-Ketoacil-Acp Precursor (PD483857; PDA1F6K2) Duas Policetídeo Beta-Ketoacil 3-Oxoacil-Acíl-Carrier-Proteína I Biosíntese sendo uma policetídeo tiolase e outra de antibióticos tipo II (PD000145; PD000224). Duas Síntase transferase de policetídeo (PD736227; PD736227).
ORF17	1: Fósforo-panteínato Acp carreadora de Acíla/ Provável Orf V de Antibiótico (PD683354)
ORF18	1: Ciclase de Biosíntese Antibiótico Policetídeo / Aromatase Síntase Oxidoreductase Orf Transferase Multifuncional (PDA1I2H8).
ORF19	1: Ciclase de Síntese de Policetídeo II/ Orf Síntase de Jadi Biosíntese Similar a Ciclase de Antibiótico (PD012644).
ORF20	1: Metil-transferase Transferase O-Metil-transferase ácida de Biosíntese de lignina Caféica 3-O-Metil-transferase Família O-Metil-transferase (PD001204) / 1: Família de Transferase O-Metil-transferase Metil-transferase (PDA7O0D6) / 1: Transferase Metil-transferase O-Metil-transferase Oxyf (PDA2B2K5) /1: Transferase O-Metil-transferase Metil-transferase (PDA7H3P9).
ORF21	2: Próximos a Domínios relacionados a enterotoxina Staphylococcal, Q4II36_GIBZE 97-405 (PDA9H8N3; PDA9H8N3) /1: Precursor de Sinal Repetitivo Paralelo Beta-Helix Kinase Epimerase de Superfície celular I 5.1.3.- Relacionada a conversão do ácido beta-D-mannuronic (M) ao ácido alpha-L-guluronic (G), Mas introduz quase que exclusivamente blocos MG, produzindo um polímero que não possui capacidade de formar gel (PD014118) / 1: Secretada (PD877708).
ORF22	Não apresentou domínios no PRODOM.