LARISSA ANTELO DE SÁ

Determinação do papel estrutural que proteínas auxiliares exercem para ativação das glicosiltransferases na biossíntese de antibióticos macrolídeos

> Dissertação apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de ciências Biomédicas da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências.

São Paulo 2017

LARISSA ANTELO DE SÁ

Determinação do papel estrutural que proteínas auxiliares exercem para ativação das glicosiltransferases na biossíntese de antibióticos macrolídeos

Dissertação apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de ciências Biomédicas da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências.

Área de concentração: Microbiologia

Orientador: Prof. Dr. Marcio Vinicius Bertacine Dias

Versão Original

São Paulo 2017

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

Ficha Catalográfica elaborada pelo(a) autor(a)

Sá, Larissa Antelo de Determinação do papel estrutural que proteínas auxiliares exercem para ativação das glicosiltransferases na biossíntese de antibióticos macrolídeos / Larissa Antelo de Sá; orientador Marcio Vinicius Bertacine Dias. -- São Paulo, 2017. 77 p.
Dissertação (Mestrado)) -- Universidade de São Paulo, Instituto de Ciências Biomédicas.
1. Produtos naturais. 2. Macrolídeos. 3. Glicosiltransferase. 4. Proteína auxiliar. I. Bertacine Dias, Marcio Vinicius , orientador. II. Título. Candidato(a): Larissa Antelo de Sá

Titulo da Dissertação: Determinação do papel estrutural que proteínas auxiliares exercem para ativação das glicosiltransferases na biossíntese de antibióticos macrolídeos.

Orientador: Prof. Dr. Marcio Vinicius Bertacine Dias

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado/Tese de Doutorado, em sessão publica realizada a, considerou o(a) candidato(a):

() Aprovado(a) () Rej	orovado(a)
-------------------------	------------

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



UNIVERSIDADE DE SÃO PAULO INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira" Av. Prof. Lineu Prestes, 2415 – CEP. 05508-000 São Paulo, SP – Brasil Telefone :(55) (11) 3091-7733 - telefax : (55) (11) 3091-8405 e-mail: cep@icb.usp.br

Comissão de Ética em Pesquisa

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB Nº 713/15 referente ao projeto intitulado: "Determinação do papel estrutural que proeínas auxiliares exercem para ativação dasglicosiltransferases na biossintese de antibióticos macrolídeos" sob a responsabilidade de Larissa Antelo de Sá, foi analisado na presente data pela CEUA - COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSH-COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da lei 11.794 de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº466 de 2012.

São Paulo, 24 de fevereiro de 2015.

PROF. DR. WOTHAN TAVARES DE LIMA Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A ZANOTTO Coordenador da CEPsh - ICB/USP

AGRADECIMETNOS

Agradeço primeiramente a Deus, por todas as benções a mim concedidas, por me proteger e guiar constantemente e principalmente por colocar no meu caminho pessoas tão boas, tornando minha jornada mais leve e prazerosa.

Aos meus pais, Vidal e Wilma e minha querida irmã Melissa. Apesar da distância sempre se mostraram presente, me apoiando em cada decisão e sempre me motivando principalmente nos momentos difíceis. Sem o amor, confiança e apoio de vocês, nada disso seria possível. Dedico todas minhas conquistas à vocês.

À Nora Antelo Perez, minha querida vó. A senhora não pôde esperar, mas onde quer que esteja eu sei que você estará presente em todos os momentos importantes da minha vida.

Ao Prof. Dr. Marcio Vinicius Bertacine Dias, por ter me concedido essa oportunidade. Sou muito grata por todos os ensinamentos e por sempre me incentivar a fazer meu melhor.

Aos meus amigos Amanda, Irene, Claudiana, Mahyara, Aline, Augusto, Vinicius e Guilherme, por toda amizade ao longo dos anos, pelo apoio e por ser minha "família" aqui no Brasil.. Obrigada por estarem presentes na minha vida tornando-a mais alegre e inesquecível, nunca me deixando sentir só.

Aos meus queridos amigos do LBEA: Priscila, Fernanda, Natalia, Tábata, Catarina, João P, João R., Sair, Gerardo e Jademilson; e aos amigos vizinhos: Cristiane, Pamela e Marcelo. Obrigada pela amizade, convivência e apoio, sem vocês essa jornada não teria sido tão alegre e a aprendizagem seria muito mais difícil.

Cristiane por sempre estar disposta a me ajudar com qualquer coisa, principalmente com biologia molecular. Obrigada por compartir protocolos e valiosas dicas! E ao Jademilson, por me ajudar com a cristalografia.

À Dra. Marilda Keico Taciro e Kaique Josuel pela disponibilidade e atenção durante os ensaios com o biorreator e ao Dr. Carlos Gustavo Baptista pela disponibilidade, atenção e auxílio com a preparação do gel nativo,

À banca examinadora por se disponibilizarem em avaliar esse trabalho, contribuindo para um trabalho mais sofisticado e elaborado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro (Bolsa 1549441) no Programa de Excelência Acadêmica (PROEX), e ao Instituto de Ciências Biomédicas (ICB) pela infraestrutura necessária para o desenvolvimento deste trabalho

"Viva como se fosse morrer amanhã aprenda como se você fosse viver para sempre"

Mahatma Gandhi

RESUMO

SÁ, L. A. Determinação do papel estrutural que proteínas auxiliares exercem para ativação das glicosiltransferases na biossíntese de antibióticos macrolídeos. 2017. 77 f. Dissertação (Mestrado em Microbiologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2017.

Produtos naturais constituem uma das principais fontes de moléculas bioativas que possuem diversas aplicabilidades. Dentre os produtos naturais, policetídeos representam uma ampla classe de compostos estruturalmente diversos cuja atividade biológica, muitas vezes está relacionada com os grupos funcionais que estão ligados ao seu esqueleto central (aglicona). Os macrolídeos representam uma classe de antibiótico amplamente utilizado e são um exemplo de policetídeos cuja atividade é dependente de moléculas de açúcares. As enzimas que realizam a glicosilação de policetídeos são as glicosiltransferases, as quais apresentam uma especificidade estrita para 6-desoxiaçúcares, porém uma especificidade relaxada para açúcares não usuais e substratos aceptores. Estudar essa flexibilidade catalítica das glicosiltransferases de produtos naturais pode contribuir para a geração de novos compostos através de glicodiversificação. Esses novos compostos podem apresentar novas atividades biológicas e propriedades farmacocinéticas melhoradas. Além da especificidade relaxada, existe um pequeno grupo de glicosiltransferases que possui um comportamento peculiar no qual uma proteína auxiliar é necessária para sua atividade catalítica, por exemplo o par TylM2/TylM3 envolvidos na biossíntese do antibiótico macrolídeo tilosina em Strepromyces fradiae. Estudar a interação e as mudanças conformacionais que ocorrem durante a formação do complexo glicosiltransferase-proteína auxiliar é fundamental para entender a influência que essas as proteínas auxiliares exercem sobre as glicosiltransferases, e com isso gerar informações que possam ser úteis na aplicação de glicodiversificação. Neste trabalho foi realizado a sublonagem de genes sintéticos que codificam a glicosiltransferase TylM2 e a proteína auxiliar TylM3 e as proteínas recombinantes foram produzidas e purificadas. Além disso foram realizadas técnicas de caracterização estrutural para proteína TylM2 no qual essa glicosiltransferase parece formar um tetrâmero em solução na ausência de sua proteína auxiliar. Ensaios de cristalização com TylM2 rendeu cristais que difratam a uma resolução abaixo de 3,0Å, mesmo após tentativas de otimização, que dificultam a determinação de sua estrutura A criação de modelos teóricos por modelagem comparativa para TylM2 e TylM3 permitiu uma investigação sobre possíveis diferenças entre a TvlM2 e suas homólogas.

Palavras-chave: Produtos naturais. Macrolídeos. Glicosiltransferase. Proteína auxiliar.

ABSTRACT

SÁ, L. A. Determining the structural role that auxiliary proteins have upon activation of glycosyltransferase in the biosynthesis of macrolide antibiotics. 2017. 77 f. Master thesis (Microbiology) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2017.

Natural products compose one of the main sources of bioactive molecules that have several applications. Amongst natural products, polyketides represent a broad class of structurally diverse compounds whose biological activity is often related to the functional groups that are linked to their central skeleton (aglycone). Macrolides represent a class of widely used antibiotic and are an example of polyketides whose activity is dependent on sugar molecules. The enzymes that perform the glycosylation of polyketides are glycosyltransferases, which have a strict specificity for 6-deoxy sugars, but a relaxed specificity for unusual sugars and acceptor substrates. Studying this catalytic flexibility of glycosyltransferases of natural products may contribute to the generation of novel compounds through glycodiversification. These novel compounds may exhibit new biological activities and improved pharmacokinetic properties. In addition to the relaxed specificity, there is a small group of glycosyltransferases who have a peculiar behavior in which an auxiliary protein is required for its catalytic activity, an example of such is the TylM2 / TylM3 pair involved in the biosynthesis of the macrolide antibiotic tylosin in Strepromyces fradiae. Studying the interaction and conformational changes that occur during the formation of the glycosyltransferase-auxiliary protein complex is critical to understanding the influence that these auxiliary proteins have on glycosyltransferases, and thereby generate information that may be useful in the application of glycodiversification. In this work, synthetic genes coding the glycosyltransferase TylM2 and the auxiliary protein TylM3 was sub cloned into pET28a vectors and the recombinant proteins were produced and purified. In addition, structural characterization techniques were performed with TylM2 which appears to form a tetramer in solution in the absence of its auxiliary protein. Crystallization assays of TylM2 yielded crystals that diffracted bellow 3.0Å and presented pathologies which prevented determining of its structure, even after attempts of optimization. The creation of theoretical models by homology modelling for TylM2 and TylM3 allowed for an investigation into possible differences that make TylM2 possess a more stringent flexibility toward acceptor substrates when compared to other homologues.

Keywords: Natural products. Macrolides. Glycosyltransferase. auxiliary protein.

LISTA DE FIGURAS

Figura 1- Exemplos da diversidade estrutural de policetídeos encontrados na natureza e suas propriedades terapêuticas
Figura 2- Exemplo do policetídeo sintase (PKS) modular do tipo I mais conhecido na literatura: a biossíntese do policetídeo 6-desoxieritronolida B que dará orígem ao antibiótico macrolídeo eritromicina A
Figura 3- A síntese de açúcares-NDP a partir de intermediários da glicólise (frutose-6-fosfato e glicose-6-fosfato) e galactose
Figura 4- Exemplos de alguns D- e L-desoxihexoses encontrados em produtos naturais glicosilados
Figura 5- Os dois mecanismos de glicosilação24
Figura 6- Estruturas secundárias de glicosiltransferases que representam os diferentes tipos de enovelamento
Figura 7- Exemplos de macrolídeos com anel macrocíclico composto por 14 membros: (A) eritromicina, 15 membros:(B) azitromicina e 16 membros: (C) tilosina
Figura 8- Estrutura da enzima EryCIII em complexo com EryCII evidenciando a formação de um hetero tetrâmero, formado por dois hetero dímeros
Figura 9- Digestão dos plasmídeos pET28a-tylM2 e pET28a-tylM3
Figura 10- Análise do teste de expressão para as proteínas TylM2 e TylM348
Figura 11- Análise da proteína TylM2 (50 kDa) produzida em <i>shaker</i> (a) e em bioreator (b).
Figura 12- Análise da proteína TylM3 (46 kDa) produzida em <i>shaker</i> (a) e em bioreator (b)
Figura 13- Curva da densidade celular das cepas BL21(DE3)pGRO7 e Tuner(DE3), utilizados para expressão das proteínas TylM2 e TylM351
Figura 14- BN-PAGE 5-15% da proteína TylM2 (50 KDa)53
Figura 15 – Espectro de dicroísmo circular da proteína TylM2 em 50 mM Tris-hcl, 150 mM NaCl, 10% v/v glicerol, $0,1\%$ v/v Thesit e 5 mM β -Mercaptoetanol, pH 7,854
Figura 16- As curvas de Guinier (a) e Krakty (b) obtidas a partir dos dados de SAXS da glicosiltransferase TylM2
Figura 17- Primeiro monocristal obtido da proteína TylM2 na condição A5- JBScreen Basic HTS (JenaBioscience) composta por 3,4 M 1,6 Hexanediol, 200 mM MgCl ₂ e 100 mM Tris-HCl pH 8,5
Figura 18- Imagem de novos monocristais da proteína TylM2 na condição A5- JBScreen Basic HTS (JenaBioscience) composta por 3,4 M 1,6 Hexanediol, 200 mM MgCl ₂ e 100 mM Tris-HCl pH 8,5
Figura 19- Imagem de cristais obtidos com os aditivos 50% v/v polietilenoglicol 400 e 40% v/v polipropilenoglicol P400, dos poços G4 e G8, respectivamente60

Figura 27- Região de interface entre TylM2 (*cartoon* em verde) e TylM3 (superfície eletrostática), com uma visão lateral (a) e frontal com (b) e sem TylM2 (c)......69

Figura 29- Interação entre a glicosiltransferase TylM2 (verde) e a proteína auxiliar TylM3 (amarelo) evidenciando os resíduos da glicosiltransferase envolvidos na interação......70

Figura 30- Interação entre a glicosiltransferase TylM2 (verde) e a proteína auxiliar TylM3 (amarelo) evidenciando os resíduos da proteína auxiliar envolvidos na interação70

LISTA DE TABELAS

Tabela 1- Reação de digestão dupla utilizando as enzimas de restrição NdeI e XhoI35
Tabela 2- Reação de ligação do inserto no vetor de expressão
Tabela 3- Condições testadas no teste de expressão
Tabela 4- Composição do meio ECPM1 e solução traços
Tabela 5- Composição de géis de empacotamento e resolução para SDS-PAGE40
Tabela 6- Composição dos géis de empacotamento e de resolução para BN-PAGE41
Tabela 7- Composição dos tampões para BN-PAGE41
Tabela 8- Resultado de espalhamento de luz dinâmico da proteína TylM2
Tabela 9- Dados de difração de raios X obtidos para TylM2 no Laboratório Nacional de LuzSincrotron
Tabela 10- Dados de difração de raios-X obtidos para TylM2 no Stanford Synchrotron Light Source
Tabela 11- Proteínas com estruturas resolvidas que apresentam similaridade à TylM262
Tabela 12- Proteínas com estruturas resolvidas que apresentam similaridade à TylM362
Tabela 13- Valores estatísticos do gráfico de Ramachandran, obtidos pelo programa PROCHECK 64

LISTA DE ABREVIATURAS

Å	Angstrom
°C	Graus célcius
μg	Micrograma (10 ⁻⁶ g)
μL	Microlitro (10 ⁻⁶ L)
μΜ	Micromolar (10 ⁻⁶ M)
APS	Persulfato de amônio
BLAST	Basic Local Alignment Search Tool
BN-PAGE	Blue Native Polyacrylamide Gel Eletrophoresis
CD	Circular Dichroism (dicroísmo circular)
cm	centímetro
CV	volume de coluna
DLS	Dynamic Light Scattering (espalhamento de luz dinâmico)
EDTA	Ácido etilenodiaminotetracético
FPLC	Fast Pressure Liquid Chromatography
g	aceleração de gravidade
g	Grama
IPTG	isopropil-β-D-tiogalactopiranosídeo
KDa	QuiloDalton (10 ³ Dalton)
L	Litro
М	Molar (mol/L)
MDa	MegaDalton
mg	miligrama (10 ⁻³ g)
mL	Mililitro $(10^{-3} L)$

mm	milímetro
mM	Milimolar(10 ⁻³ M)
ng	Nanograma (10 ⁻⁹ g)
nm	nanometro
pb	pares de base
PDB	Protein Data Bank
PMSF	Fluoreto de fenilmetano sulfonilo
Rg	Raio de giro
rpm	rotação por minuto
SAXS	Small Angle Light Scattering (espalhamento de luz a baixos ãngulos)
SDS	Dodecil sulfato de sódio
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Eletrophoresis
TDP	Timidina-di-fosfato
TEMED	N,N,N',N'-tetramethylenediamine
U	Unidades
V	Volts
vvm	Volume de ar por volume de meio por minuto
v/v	Volume por volume
w/v	Massa por volume

SUMÁRIO

1 INTRODUÇÃO1	.7
1.1 Produtos Naturais1	7
1.2 Biossíntese de Policetídeos1	8
1.3 Açúcares NDP1	9
1.4 Glicosiltrasnferases2	22
1.5 Antibióticos Macrolídeos2	25
1.6 Glicodiversificação2	28
2 OBJETIVOS E JUSTIFICATIVA	62
3 MATERIAIS E METODOS	64
3.1 Sub clonagem dos genes sintéticos tylm2 e tylm3	\$4
3.2 Eletroforese em Gel de Agarose	6
3.3 Teste de Expressão em Pequena Escala	6
3.4 Expressão Heteróloga de Proteínas Recombinantes	57
3.5 Expressão de Proteínas Recombinantes em Biorreator	8
3.6 Extração e Purificação das Proteínas	i9
3.7 Eletroforese de Proteínas4	0
<u>3.7.1 SDS-PAGE</u>	0
<u>3.7.2 BN-PAGE</u>	1
3.8 Dicroísmo Circular	2
3.9 Espalhamento Dinâmico de Luz4	2
3.10 Cristalização De Proteínas4	3
3.11 Coleta de Dados de Difração de Raio-X4	3
3.12 Determinação da Estrutura Cristalográfica4	3
3.13 Espalhamento de Raios-X a Baixo Ângulo4	4
3.14 Modelagem Molecular por Homologia4	4

4 RESULTADOS E DISCUSSÃO	46
4.1 Clonagem, Expressão e Purificação	46
<u>4.1.1 Sub-clonagem, Expressão e Purificação das Proteínas TylM2 e TylM3</u>	46
4.2 Caracterização Estrutural da Proteína TylM2	52
<u>4.2.1 Gel Nativo</u>	52
4.2.2 Dicroismo circular	53
4.2.3 Espalhamento de luz dinâmico	54
4.2.4 Espalhamento de raios-X a baixo ângulo	55
4.3 Cristalização, Difração e Determinação da Estrutura	56
4.3.1 Ensaios biofísicos da proteína TylM2	56
4.4 Modelagem molecular e análise de Bioinformática para TylM2 e TylM3	61
5 CONSIDERAÇÕES FINAIS	71
REFERENCIAS*	72

1 INTRODUÇÃO

1.1 Produtos Naturais

Produtos naturais, também denominados de metabólitos secundários, constituem uma das principais fontes de moléculas bioativas com propriedades terapêuticas que podem ser utilizadas na área clínica humana e veterinária. Os policetídeos representam uma das classes de metabólitos secundários que apresentam uma grande diversidade estrutural, sendo produzidos por plantas, fungos e bactérias. Além disso, apresentam uma ampla variedade de propriedades terapêuticas tais como: atividades antibióticas, antifúngicas, antitumorais, antivirais, antiparasitárias e imunossupressoras (MÉNDEZ; SALAS, 2001). A Figura 1 mostra a diversidade estrutural de alguns policetídeos e suas propriedades terapêuticas.



Figura 1- Exemplos da diversidade estrutural de policetídeos encontrados na natureza e suas propriedades terapêuticas. Modificado de: Staunton, Weissman (2001).

1.2 Biossíntese de Policetídeos

A biossíntese de policetídeos, denominada PKS, envolve uma série de reações de condensação e modificação a partir de metabólitos primários (geralmente são derivados de propionil-CoA, acetil-CoA, malonil-CoA, propianato-CoA e metilmalonil-CoA) resultando na formação de uma estrutura central denominada aglicona (HEIA et al., 2011). As reações de condensação e modificação são catalisadas pelas policetídeo sintases (PKSs do inglês PolyKetide synthase). A síntese de policetídeos em bactérias é dividida em três grupos, baseado no arranjo das policetídeo sintases: O PKS tipo I é constituído por uma ou mais proteínas compostas por unidades, que compõem os módulos (Figura 2). Cada módulo possui os três domínios principais que são necessários para catalisar as reações de extensão, sendo eles os domínios KS (cetosintase), AT (aciltransferase) e ACP (proteína carreadora de acila). Além desses três, outros domínios que catalisam reações de modificações do grupo ceto (aglicona) são encontrados, tais como os domínios KR (cetoredutase), DR (dehidratase) e ER (enoil redutase). O PKS tipo II, como a PKS tipo I, é composta por várias proteínas, porém cada uma apresenta apenas um domínio (CHAN et al., 2009). O PKS tipo III difere dos tipos I e II por não utilizar um domínio ACP (proteína carreadora de acila) e se assemelham à PKS de plantas por apresentarem domínios relacionados à família de policetídeos CHS (calcona sintase) e STS (estilbena sintase) (STAUNTON; WEISSMAN, 2001).

A grande variedade funcional e estrutural dos policetídeos está relacionada aos seguintes fatores: diversidade de unidades iniciadoras e de extensão; tamanho da cadeia formada, que é determinado pelo número de ciclos de extensão; grau de oxidação dos grupos ceto e a forma pela qual a síntese do policetídeo é terminada, podendo ser por ciclização ou hidrólise (MOORE; HERTWECK, 2002). Além disso, ocorrem outras modificações póssíntese do policetídeo que aumenta ainda mais sua variabilidade, tais como: oxidação, no qual os grupos hidroxila e carbonila são adicionadas ao aglicona; metilação dos átomos de carbono, oxigênio e nitrogênio e glicosilação, onde ocorre a decoração do aglicona com uma ou mais moléculas de açúcares-NDP (STAUNTON; WEISSMAN, 2001).



Figura 2- Exemplo do policetídeo sintase (PKS) modular do tipo I mais conhecido na literatura: a biossíntese do policetídeo 6-desoxieritronolida B que dará orígem ao antibiótico macrolídeo eritromicina A. Modificado de: Cane (2010).

No grupo PKS do tipo I encontramos alguns policetídeos que apresentam aplicabilidade terapêutica por possuírem ação antibacteriana, tais como tilosina, eritromicina, monensina A, rifamicina, cervimicina entre outros (STAUNTON; WEISSMAN, 2001). Esses policetídeos são compostos glicosilados, ou seja, possuem em sua estrutura um ou mais moléculas de açúcares os quais foram adicionados à aglicona por reações de glicosilação, que consiste na transferência de uma molécula de açúcar a partir de uma molécula doadora (açúcar ativado) para outra aceptora (aglicona) por meio de enzimas denominadas glicosiltransferases. Os açúcares transferidos geralmente são hexoses e muitas vezes a sua presença nos produtos naturais é essencial para que o composto apresente atividade biológica (THIBODEAUX; MELANÇON; LIU, 2008). Na Figura 2, a reação de glicosilação é realizada pelas enzimas EryBV e pelo par EryCIII/EryCII, as quais adicionam L-micarose e D-desosamina, respectivamente ao aglicona 6-desoxieritronalida B.

1.3 Açúcares NDP

Galactose e os intermediários da glicólise (glicose-6-fosfato e frutose-6-fosfato) são as principais fontes de carbono e energia utilizadas pelos seres vivos e além disso são moléculas precursoras utilizadas na glicosilação de metabólitos secundários, porém precisam ser ativados à açúcar-NMP (nucleotídeo monofosfato) ou açúcar–NDP (nucleotídeo difosfato).

Nessa forma ativada do açúcar, o grupo fosfonucleotídeo serve como elemento de reconhecimento para enzimas biossintéticas durante a reação de transferência do açúcar, além de fornecer a energia necessária para a reação (THIBODEAUX; MELANÇON; LIU, 2008). A Figura 3 mostra a base biosintética dos açúcares-NDP mais encontrados. Na qual a frutose-6-fosfato é convertida em manose-6-fosfato por ação da enzima fosfo-mano-isomerase durante a biossíntese de açúcares-GDP, e em glicosamina-6-fosfato pela glicosamina-6-fosfato e açúcares-UDP. A glicose-6-fosfato é também um precursor biossintético de açúcares-UDP, porém é mais comum que seja utilizado para a biossíntese de açúcares-TDP e –CDP. Outra forma de se obter açúcares-UDP é pelo catabolismo da galactose pela via de Leloir. Em todos os casos, o açúcar-6-fosfato é convertido a açúcar-1-fosfato por fosfohexose mutases antes de ocorrer a nucleotidilação, que consiste na transferência do grupo fosfonucleotídeo às moléuculas de açúcar realizado pelas nucleotidil transferases. (THIBODEAUX; MELANÇON; LIU, 2008).



Figura 3- A síntese de açúcares-NDP a partir de intermediários da glicólise (frutose-6-fosfato e glicose-6-fosfato) e galactose. Esses monossacarídeos-6-fosfatos precisam convertidos a monossacarídeos-1-fosfato para serem ativados pelos nucleotidiltransferases (uridilil, guanadilil, timidilil, citidilil para formarem açúcares-UDP,-GDP, - TDP e -CDP, respectivamente).

Modificado de: Thibodeaux, Melaçon, Liu (2008).

Elshahawi et al. (2015) relatam a existência de 15.940 produtos naturais de origem bacteriana sendo que destes, 12.514 representam produtos naturais não glicosilados e 3426 representam aqueles glicosilados. Ademais, foram contabilizados 344 tipos de moléculas de açúcares não usuais encontrados em produtos naturais bacterianos (ELSHAHAWI et al., 2015). Essa grande variedade de açúcares doadores está atribuída aos diferentes substituintes e a estereoquímica dos átomos de carbonos 2-5. Além disso, podem ocorrer mais reações de desoxigenações, transaminações e metilações nos átomos de carbono, nitrogênio e oxigênio na molécula do açúcar doador, dando origem a desoxi, amino e metoxi-açúcares (MÉNDEZ; SALAS, 2001).

Os açúcares-TDP compõem a classe mais estruturalmente diversificada que é encontrada na natureza, sendo o açúcar doador mais usual para a biossíntese de produtos naturais glicosilados (THIBODEAUX; MELANÇON; LIU, 2008). A maioria dos açúcares TDP são desoxiaçúcares sendo desoxigenados no carbono 2, 3 ou 4 do anel piranose. A biossíntese de açúcares TDP ocorre a partir de uma molécula de glicose-1-fosfato, que é convertida em TDP-D-glicose pela enzima timidililtransferase e posteriormente convertido a TDP-4-ceto-6-desoxi-D-glicose pela enzima TDP-D-glicose 4,6 desidratase. A conformação D ou L dos açúcares é resultado da 5- ou 3,5-epimerase (MÉNDEZ; SALAS, 2001). Os principais D- e L-desoxihexoses podem ser vistos na figura 4 abaixo. Como TDP-4-ceto-6-desoxi-D-glicose é um intermediário chave na biossíntese de desoxi-açúcares em bactérias, a maioria dos clusters biossintéticos de produtos naturais possuem genes que codificam as enzimas timidililtransferase e 4,6-desidratase (THIBODEAUX; MELANÇON; LIU, 2008).



Figura 4- Exemplos de alguns D- e L-desoxihexoses encontrados em produtos naturais glicosilados. Fonte: Méndez, Salas (2001).

1.4 Glicosiltrasnferases

As glicosiltransferases pertencem a um grupo de enzimas cuja função é catalisar a transferência de uma molécula de açúcar a partir de uma molécula doadora à uma molécula aceptora. O sítio ativo da glicosiltransferase fornece um microambiente para as moléculas doadoras e aceptoras se ligarem e para que ocorra a reação de glicosilação (LIANG et al., 2015). De acordo com o banco de dados de enzimas de carboidrato ativo (CAZy – disponível em: http://www.cazy.org) existem 236.104 glicosiltransferases, conhecidas ou preditas, classificadas em 98 famílias de acordo com similaridade na sequência de aminoácidos. Glicosiltransferases podem ser ainda classificadas de acordo com o mecanismo de glicosilação e com o enovelamento (WILLIAMS; THORSON, 2009). Apesar da grande diversidade na sequência de aminoácidos devido a processos de evolução, não há uma grande

diversidade quanto ao mecanismo de glicosilação. Desta maneira são encontrados apenas dois mecanismos principais: retenção e inversão; Por outro lado quanto aos tipos de enovelamento, são encontrados cinco famílias de glicosiltransferases: GT-A, GT-B, GT-C, GT-D e GT-E (LIANG et al., 2015).

Para a maioria das glicosiltransferases caracterizadas, o substrato doador é um açúcar-NDP ativado, entretanto, -NMP e poliproprenil difosfato também podem servir como açúcares doadores para determinadas glicosiltransferases (THIBODEAUX; MELANÇON; LIU, 2008). O substrato aceptor para as glicosiltransferases relacionadas com a biossíntese de produtos naturais são estruturalmente diversificados e inclui várias classes de compostos que são: peptídeos não ribossomais, por exemplo a vancomicina e teicoplanina; aminocoumarinas como noviobiocina e couremicina; enedienos como caliqueamicina; macrolactonas como vicenistatina; policetídeos aromáticos como antraciclinas e macrolídeos como eritromicina e tilosina (WILLIAMS; THORSON, 2009).

A reação de glicosilação constitui no deslocamento de um substituinte anomérico do açúcar doador por um grupo funcional nucleofílico da molécula aceptora, que pode ser a aglicona ou até mesmo outros açúcares, formando assim uma ligação glicosídica (THIBODEAUX; MELANÇON; LIU, 2008). A ligação glicosídica mais comum é aquela realizada entre o substituinte anomérico do açúcar doador com o oxigênio nucleofílico de uma hidroxila da molécula aceptora, formando uma ligação O-glicosídica. Porém a ligação pode envolver átomos de nitrogênio, formando ligações N-glicosídicas; enxofre, formando a ligações S-glicosídicas e carbono, formando ligações C-glicosídicas (LAIRSON et al., 2008). Quando se trata de produtos naturais glicosilados, a ligação glicosídica mais comum é as O-glicosídicas (LIANG et al., 2015).

Como já citado acima, existem dois mecanismos de glicosilação, os quais estão relacionados com a estereoquímica da ligação glicosídica formada (Figura 5). No mecanismo de inversão, as glicosiltransferases catalisam a transferência da molécula de açúcar e a ligação glicosídica é formada com a inversão da configuração do carbono anomérico do açúcar enquanto que no mecanismo de retenção, a ligação glicosídica é formada e a configuração do carbono anomérico é mantida (WILLIAMS; THORSON, 2009).



Figura 5- Os dois mecanismos de glicosilação. À esquerda o mecanismo de retenção no qual a configuração do carbono anomérico após a formação da ligação glicosídica entre o açúcar e a molécula aceptora (aglicona) é mantida, e à direita, o mecanismo de inversão mostrando a mudança da configuração do carbono anomérico.

Na classificação enovelamento glicosiltransferases, baseada no de cinco enovelamentos típicos compõem as famílias GT-A, GT-B, GT-C, GT-D e GT-E. A família GT-A (Figura 6a) é caracterizada por um único domínio com a topologia α,β,α que se assemelha a um enovelamento do tipo Rossman (THIBODEAUX; MELANÇON; LIU, 2008). A folha β central composta por 7 fitas β é flanqueada por α hélices e outra folha β antiparalela menor. Na associação entre as duas folhas β , região onde o açúcar- NDP se liga, é encontrado o motivo conservado DXD (Asp-X-Asp) no qual um íon metal bivalente irá se ligar, comumente Mn²⁺ ou Mg²⁺, servindo como ponto de ancoragem para um grupo difosfato do açúcar NDP (LIANG et al., 2015).

Representantes da família GT-B (Figura 6b) apresentam dois domínios separados, unidos entre si por um *linker*. Cada domínio também se assemelha ao enovelamento tipo *Rossman* sendo composto por folhas β paralelas ligada à α hélices. Entre os dois domínios ocorre a formação de uma fenda profunda, que é o local onde as moléculas aceptoras e doadoras se ligam (LAIRSON et al., 2008). Glicosiltransferases que pertencem a essa família geralmente não dependem de íons de metais bivalentes e não apresentam um motivo conservado. A ligação da molécula aceptora e o açúcar doador ocorrem no domínio N- e Cterminal, respectivamente, sendo que o domínio C-terminal é mais conservado no qual uma região rica em resíduos de glicina é encontrada (LIANG et al., 2015), enquanto que o domínio N- terminal é menos conservado, apresentando algumas variações nos loops e α -hélices próximos ao sítio ativo. Essas variações podem ser fruto da evolução para que estas enzimas pudessem se ligar a diferentes substratos. Acredita-se que a maioria das glicosiltransferases relacionadas com a biossíntese de produtos naturais bacterianos pertence à família GT-B (BRETON; et al., 2006).

A família GT-C (Figura 6c) possui um domínio N-terminal que é um domínio trans membrana e o domínio C-terminal, composto por um conjunto de α hélices que apresenta atividade de glicosiltransferase (LIANG et al., 2015). Representantes desse grupo utilizam açúcares ligados a lipídeos como moléculas doadores e a glicosilação é dependente de um cátion bivalente (GLOSTER, 2014).

Recentemente a estrutura de uma enzima de função desconhecida denominada DUF1792 foi determinada, evidenciando que esse composto era uma glicosiltransferase envolvida na biossíntese de O-glicanos em bactérias. Por apresentar um enovelamento diferente das famílias GT-A. GT-B e GT-C, foi designado como a nova família GT-D (Figura 6d). Nessa família há um motivo altamente conservado DxE (Asp-X-Glu) (ZHANG et al., 2014). Os representantes da família GT-51 são chamadas de "tipo lisozima", pois suas estruturas secundárias são bastante semelhantes à lisozima de bacteriófagos λ , apesar de não possuir nenhuma homologia sequencial (Figura 6e) (LIANG et al., 2015).



Figura 6- Estruturas secundárias de glicosiltransferases que representam os diferentes tipos de enovelamento. (a) GT-A: SpsA (PDB: 1QGS) da família GT2 de Bacillus subtilis; (b) GT-B: SpnG (PDB: 3TSA) da família GT1 de Saccharopolyspora spinosa; (c) GT-C: PglB (PDB: 3RCE) da família GT66 de *Campylobacter lari*; (d) GT-D: DUF1792 (PDB: 4PFX) da família GT101 de *Streptococci parasanguini* e (e) GT-tipo lisozima: PGT (PDB: 2OQO) da família GT51 de *Aquifez aeolicus*. As regiões em azul representam regiões envolvidas na glicosilação, α hélices estão em vermelho, folhas β em amarelo e loops em verde.

1.5 Antibióticos Macrolídeos

Dentre os antibióticos que apresentam sua atividade dependente de moléculas de açúcares temos os macrolídeos, que constituem uma das primeiras classes de agentes antimicrobianos descoberta (BURGIE; HOLDEN, 2008; GOLDMAN; SCAGLIONE, 2004). A palavra macrolídeo foi inicialmente proposta por Robert Woodward em 1954 para descrever produtos naturais que continham um anel lactona macrocíclica composto por 12 a 16 átomos de carbono ligados a uma ou mais moléculas de açúcar. Porém, atualmente o termo macrolídeo abrange compostos que possuem um anel de até 62 átomos e também os semissintéticos que possuem átomos de nitrogênio no anel (figura 7). Mais de 2000 macrolídeos produzidos por bactérias, fungos, liquens, algas, plantas, invertebrados e vertebrados foram descritos, sendo que aproximadamente 40% são produzidos unicamente pela classe actinomicetos, sobretudo pelo gênero *Streptomyces* (OMURA, 2002).



Figura 7- Exemplos de macrolídeos com anel macrocíclico composto por 14 membros: (A) eritromicina, 15 membros:(B) azitromicina e 16 membros: (C) tilosina.
Fonte: Dinos; Michelinaki; Kalpaxis (2001).

Os antibióticos macrolídeos apresentam um mecanismo de ação no qual o alvo é a tradução proteica (BORISOVA et al., 2006). Estudos estruturais realizados por Harms et al. (2001) e Hansen et al. (2002) revelaram informações estruturais sobre a interação de macrolídeos de 14 átomos com a subunidade ribossomal 50S de *Deinococcus radiodurans*, e de macrolídeos de 15 e 16 átomos em complexo com a subunidade ribossomal 50s de *Haloarcula marismortui*, respectivamente. Esses estudos revelaram que os macrolídeos se ligam ao túnel de saída da cadeia polipeptídica em crescimento na subunidade 50S, predominantemente através do contato com rRNA 23S. Além disso, é através das moléculas de açúcar que é realizado o contato do antibiótico com o alvo (HANSEN et al., 2002). Essa

interação com a subunidade ribossomal maior é o que torna essa classe de antibiótico altamente seletivo e, portanto muito utilizado na terapia clínica. Geralmente os macrolídeos são ativos principalmente contra bactérias gram-positivas, embora apresentem uma limitada ação contra bactérias gram-negativas (OMURA, 2002).

Geralmente, os macrolídeos são utilizados para o tratamento de infecções causadas por *Mycoplasma pneumoniae*, doença dos legionários, infecções por clamídia, difteria, coqueluche, infecções estreptocócicas, infecções estafilocócicas, infecções por diversas bactérias, incluindo *Campylobacter*, *Helicobacter pylori*, *Mycobacterium tuberculosis* entre outras (BRUNTON; LAZO; PARK, 2010). Interessantemente alguns macrolídeos como a azitromicina também apresenta efeitos anti-inflamatórios, inibindo a produção de algumas interleucinas e assim, reduzindo a inflamação neutrofílica pulmonar, e portanto, podem ser utilizados no tratamento de doenças inflamatórias, tais como fibrose cística, bronquiolite obliterante, asma e bronquiolite aguda (LUISI et al., 2012). Estudos recentes demonstraram que dois compostos macrolídeos de 36 átomos produzidos pelo micro-organismo marinho *Streptomyces caniferus* apresenta ação citotóxica contra células humanas de carcinoma de pulmão, adenocarcinoma da mama e carcinoma colorretal, juntamente com uma ação antifúngica fraca contra *Candida albicans* (PÉREZ et al., 2015).

Por outro lado, existem duas classes de resistência à antibióticos, a resistência intrínseca e a resistência adquirida. A resistência intrínseca está relacionada com características inerentes do microrganismo e faz com que o antibiótico não possua atividade enquanto que a resistência adquirida pode surgir a partir de mutações em determinados genes ou ser adquirido por plasmídeos ou transposons que podem ser transferidos horizontalmente de uma bactéria à outra. Além disso, o uso contínuo e indiscriminado de antibióticos gera uma pressão seletiva fazendo com que resistência a antibióticos seja adquirida (OMURA, 2002).

Vários mecanismos de resistência têm sido observados para essa classe de antibiótico, sendo eles: 1) a bomba de efluxo mediado pelos genes *msr*, *mef* e *mre*; 2) modificações no alvo do antibiótico mediado pelo gene *erm*, e 3) a inativação do antibiótico mediado pelo genes: *ere* codificando a eritromicina esterase, o gene *vgb* que codifica estreptogramina B hidrolase, macrolídeo fosfotransferase codificada pelo gene *mph* e lincosamina nucleotidiltransferase codificado pelo gene *lin* (OMURA, 2002). A resistência intrínseca encontrada em todos procariotos é a bomba de efluxo, uma maquinaria dependente de energia

utilizada para transportar moléculas estranhas e tóxicas do interior da célula para o meio externo sendo impedidos de atingir níveis intracelulares suficientes para ter a ação esperada (WALSH, 2000). Bombas de efluxo podem ser específicas para apenas um tipo de substrato ou podem ser capazes de reconhecer compostos estruturalmente diversos. Em procariotos são encontrados cinco tipos de bombas associadas com resistência à antibióticos e todos utilizam a força próton motrix como fonte energética, com exceção de transportadores ABC, (WEBBER; PIDDOCK, 2003). Na modificação do alvo, pode ocorrer a metilação de um resíduo de adenina, A2058 na unidade 23S realizada pela metiltransferase Erm utilizando S-adenosilmetionina (SAM) como co-fator (WALSH, 2000). A inativação do macrolídeo pode ocorrer através de modificações enzimáticas, por exemplo algumas cepas de *Escherichia coli* patogênicas produzem enzimas que clivam o anel macrocíclico. Ainda, foi detectada atividade de glicosiltransferase em organismos não produtores de macrolídeos que demonstraram ser capazes de inativar o antibiótico macrolídeo através da glicosilação do antibiótico, tornando-o inativo (CUNDLIFFE, 1992).

1.6 Glicodiversificação

Devido à necessidade e a importância do desenvolvimento de novos derivados de antibióticos, várias abordagens são utilizadas com o objetivo de modificar antibióticos a partir de fontes naturais visando alterar suas especificidades ou diminuir sua toxidade. A glicodiversificação é uma técnica utilizada para se obter diferentes derivados de antibióticos através da adição ou eliminação de açúcares no esqueleto central do antibiótico (THIBODEAUX; MELANÇON; LIU, 2007). Podem-se utilizar duas estratégias para a glicodiversificação. A primeira é geração de novas estruturas não usuais através de modificações químicas de grupos funcionais de um açúcar precursor e a segunda é a produção de diferentes glicoformas pela alteração da adição de diferentes açúcares a uma molécula aceptora.

Com o uso das glicosiltransferases, utilizam-se técnicas *in vivo*, na qual se emprega a recombinação gênica para deletar, substituir, ativar ou acrescentar genes em uma determinada espécie produtora de antibiótico, como já realizado com sucesso para os antibióticos macrolídeos eritromicina e pikromicina (JUNG et al., 2007; THIBODEAUX; MELANÇON; LIU, 2007; WU et al., 2016); e técnicas *in vitro*, que consiste na incubação da aglicona com diferentes doadores utilizando a enzima glicosiltransferase purificada, como realizado por Borisova et al. (2006) para investigar a especificidade do par DesVII/DesVIII Existem ainda

as técnicas na qual os aminoácidos do sítio ativo são mutados (THIBODEAUX; MELANÇON; LIU, 2007) ou a troca de domínios funcionais entre diferentes glicosiltransferases, gerando proteínas quimeras que podem aceitar tanto aceptores quanto doadores diferentes (TRUMAN et al., 2009). Várias pesquisas revelam que, as glicosiltransferases apresentam uma alta especificidade para 6-desoxi açúcares porém uma especificidade relaxada tanto para as moléculas doadoras, quanto para os aceptores. Portanto, explorando a flexibilidade catalítica dessas enzimas é possível encontrar alternativas para a criação de novos derivados de macrolídeos (BORISOVA et al., 2008). Xue et al. (1998) sugerem, por exemplo, que a glicosiltransferase codificada pelo gene *desVII*, é capaz de catalisar a glicosilação tanto de macrolactonas de 12 quanto de 14 átomos.

Liang et al. (2015) relatam que atualmente são conhecidas onze glicosiltransferases que requerem a presença de proteínas auxiliares para glicosilação de produtos naturais: DesVII/DesVIII relacionadas com a biossíntese de pikromicina/metimicina em Streptomyces venezuelae; EryCIII/EryCII relacionada com a síntese de eritromicina em Saccharopolyspora erythrea; TylM2/TylM3 relacionadas com a síntese de tilosina em Streptomyces fradiae; MycB/MydC, relacionadas com a síntese de micinamicina em Micromonospora griseorubida; MegCIII/MegCII e MegDI/MegDVI relacionadas com a síntese de megalosamina em Micromonospora megalomicea; AngMII/AngMIII relacionadas com a síntese de angolamicina em Streptomyces eurythermus; Srm5/Srm6 e Srm29/Srm28 relacionadas com a síntese de spiramicina em Streptomyces ambofaciens; PnxGT1/PnxO5 relacionado com a síntese de FD-594 em Streptomyces sp. TA-0256 e AknS/AknT relacionados com a síntese de aclacinomicina A em Streptomyces galilaeus (LIANG et al., 2015). Apesar da alta homologia sequencial com o citocromo P450, as proteínas auxiliares não possuem o resíduo de cisteína altamente conservado em citocromos P450 que está relacionado com a ligação do grupo heme. Além disso, as proteínas auxiliares apresentam uma longa α-hélice no amino terminal da proteína que parece ter um papel importante na ativação de suas respectivas glicosiltransferases (ISIORHO et al., 2014).

A dependência de glicosiltransferase por uma proteína auxiliar foi primeiro demonstrada *in vitro* para a glicosiltranferase DesVII, que catalisa a ligação do TDP-D-desosamina ao 10-desoximetilonida ou narbonolida na biossíntese de metimicina, neometimicina, narbomicina e pikromicina em *Streptomyces venezuelae* (BORISOVA et al., 2004; BORISOVA; LIU, 2010). A caracterização de metabólitos produzidos por mutantes de

Streptomyces venezuelae com deleção do gene *desVIII* mostrou que seu produto, uma proteína auxiliar, é importante para a biossíntese de macrolídeos glicosilados, aumentando significativamente a reação de glicosilação da aglicona pela glicosiltransferase DesVII (BORISOVA; LIU, 2010). Estudos de complementação interespécie mostraram que DesVII consegue utilizar açúcares exógenos para glicosilar as agliconas nativas 10-desoximetinolida e narbonolida (BORISOVA et al., 2006; HONG et al., 2005) e Jung et al. (2007) mostraram que DesVII/DesVIII é capaz de glicosilar agliconas não nativas, como a tilactona de 16 membros. Além disso, quando realizado a troca da proteína auxiliar (DesVIII) da glicosiltransferase DesVII com a proteína auxiliar (TylM3) da glicosiltransferase TylM2, o par DesVII/TylM3 não conseguiu glicosilar agliconas de 12, 14 e 16 membros (10-desoximetilonida, naronolida e tilactona, respectivamente), o que foi também observado para o par TylM2/DesVIII. Porém, trocando o par DesVII/DesVIII por TylM2/TylM3, somente a tilactona foi glicosilada, sugerindo que a glicosiltransferase TylM2 apresenta uma especificidade ao substrato aceptor (aglicona) muito mais rígida quando comparada à DesVII. (JUNG et al., 2007).

Um estudo mais recente realizado por WU et al. (2016) visou investigar *in vivo* essa flexibilidade do par DesVII/DesVIII em *Saccharopolyspora erythrea*. Esses estudos sugerem que EryCII possivelmente ativa a glicosiltransferase DesVII *in vivo* formando o complexo heterólogo DesVII/EryCII, entretanto este apresenta uma atividade catalítica baixa quando comparado com os pares nativos. Possivelmente os pares nativos possuem uma maior afinidade um ao outro e acredita-se que há a formação de um complexo estável *in vivo* (WU et al., 2016), como mostrado na estrutura tridimensional obtido pelo complexo EryCII/EryCIII (Figura 8) (MONCRIEFFE et al., 2012). A função da proteína auxiliar assemelha-se a de uma chaperona que permite o correto enovelamento da glicosiltransferase, atuando como um ativador alostérico (MONCRIEFFE et al., 2012; YUAN et al., 2005). Porém não é claro se o complexo, uma vez formado é mantido durante a glicosilação com a formação de um trímero de heterodímero para DesVII e DesVIII (BORISOVA; LIU, 2010); ou se após remoção da proteína auxiliar, a glicosiltransferase ainda se mantém "ativada", como demonstrado com



Figura 8- Estrutura da enzima EryCIII em complexo com EryCII evidenciando a formação de um hetero tetrâmero, formado por dois hetero dímeros.
Fonte: Moncrieffe et al. (2012).

Através dos dados estruturais, Moncrieffe et al. (2012) pôde inferir que a proteína acessória EryCII, que apresenta alta similaridade com o citocromo P450, liga-se quase que exclusivamente no domínio N-terminal da glicosiltranferase EryCIII apesar de não apresentar interações diretas. Embora a estrutura obtida do complexo mostre a interface de interação entre estas duas proteínas, poucos detalhes puderam ser obtidos em relação ao sítio ativo, devido à baixa resolução. Deste modo, o melhor entendimento desta classe de glicosiltransferase só poderá ser obtido com a determinação da estrutura de outras glicosiltransferases similares a partir de uma resolução com maiores detalhes estruturais.

2 OBJETIVOS E JUSTIFICATIVA

Macrolídeos representam uma larga classe de antibióticos amplamente utilizados na terapia clínica. Entretanto o uso indiscriminado de antibióticos, em geral, tem resultado no surgimento de cepas resistentes à essas moléculas bioativas. Uma das alternativas de obter novas moléculas capazes de superar os mecanismos de resistência seria o desenvolvimento de novas moléculas com modificações em seus padrões de glicosilações, uma vez que, os grupos de açúcares são indispensáveis para que estas moléculas apresentem bioatividade. Glicosiltransferases envolvidas na biossíntese de antibióticos apresentam uma especificidade mais flexível ao substrato doador, o que os tornam uma ferramenta atraente para serem utilizadas em técnicas de glicodiversificação, que visa transferir diferentes moléculas de açucares à diferentes posições. Entretanto, para a aplicação desta técnica, o conhecimento da estrutura tridimensional e do seu mecanismo catalítico torna-se crucial.

Apesar de Moncrieffe et al. (2012) ter determinado a estrutura do primeiro complexo entre as enzimas EryCII-EryCIII, a resolução de 3.1 Å não permitiu propor de forma confiável um mecanismo catalítico de como ocorre essa interação. Além disso, somente o complexo entre a glicosiltransferase e a proteína auxiliar foi obtido, não evidenciando as reais alterações ou efeitos alostéricos os quais a proteína auxiliar causa na glicosiltransferase. Assim, existe uma lacuna muito grande de investigações estruturais a respeito desta classe de glicosiltransferases que necessitam ser estudadas. A obtenção destas informações estruturais ajudará compreender essa classe de enzimas, bem como gerará informações úteis que poderão ser utilizadas na aplicação da biologia sintética para produção de novos derivados de macrolídeos. Dentro desse contexto, esse trabalho visa caracterizar a glicosiltransferase TylM2 e sua proteína auxiliar TylM3 relacionada com a biossíntese do antibiótico macrolídeo tilosina e têm como objetivos:

- O desenvolvimento de técnicas para produção das proteínas de interesse com o intuito de realizar ensaios de caracterização estrutural e biofísico. Essas técnicas incluem: clonagem dos genes sintéticos *tylm2* e *tylm3* no vetor de superexpressão pET28a transformados em cepas de *Escherichia coli* BL21(DE3), seguida da expressão e purificação das proteínas de interesse.
- A caracterização da proteína TylM2 em solução para determinação da composição de sua estrutura secundária e do seu estado oligomérico através das técnicas de dicroísmo

circular e espalhamento de luz dinâmico, gel nativo, e espalhamento de luz a baixos ângulos.

- Identificação e otimização de condições de cristalização da proteína TylM2 para a obtenção de cristais que, por difração de raios-X e determinação de sua estrutura tridimensional poderá gerar informações a respeito de mudanças geradas pela interação entre a glicosiltransferase e sua respectiva proteína auxiliar.
- Aplicar técnicas de bioinformática e modelagem molecular para a construção de modelos tridimensionais por homologia da glicosiltransferase (TylM2) e a proteína auxiliar (TylM3), os quais poderão ser utilizados para investigação estrutural das duas proteínas.

3 MATERIAIS E METODOS

3.1 Sub clonagem dos genes sintéticos tylm2 e tylm3

Os genes sintéticos, adquiridos junto a empresa GenScript, foram otimizados para expressão em *Escherichia coli* apresentando um balanceamento de conteúdo CG e a eliminação de sítios de restrição internos, presente nos genes selvagens referentes às glicosiltransferases e sua proteína acessória envolvida na via biossintética do antibiótico macrolídeo tilosina: *tylM2* e *tylM3*, respectivamente. Ambos os genes foram adquiridos clonados no vetor pUC57 com sítios para as enzimas de restrição *NdeI* (CATATG) e *XhoI* (CTCGAG).

Esses plasmídeos foram inseridos em células de *Escherichia coli* DH5 α através de reações de transformação por choque térmico. Para cada reação foram utilizados 50 µL de células competentes de *E. coli* DH5 α que foram incubados em gelo juntamente com 150 ng de DNA plasmidial por 5 minutos. Em seguida a reação foi incubada a 42 °C por 40 segundos para o choque térmico e posteriormente incubado em gelo novamente por mais 10 minutos. Em seguida ao choque térmico 100 µL de meio LB (10 g NaCl, 10 g triptona e 5 g extrato de levedura) foram adicionados à reação e incubado a 37 °C sob agitação constante de 200 rpm. Após 1 hora, a reação foi espalhada em placa de petri contendo 20 mL de meio LB sólido acrescido de 50 µg/mL do antibiótico seletivo ampicilina e em seguida foi incubado a 37 °C por 18 horas.

Algumas colônias crescidas na placa de petri foram selecionadas aleatoriamente, as quais foram inoculadas em culturas de 10 mL de meio LB contendo 50 μ g/mL de ampicilina e incubados a 37 °C sob agitação constante de 200 rpm por 18 horas. Em seguida foi realizada a extração de DNA plasmidial presente nas células. Para isso, as culturas foram centrifugadas a 12,000 *g* por 10 minutos, com sucessivos passos descritos no protocolo estabelecido pelo fabricante GeneJET Plasmid Miniprep Kit (ThermoFischer Scientific). As reações provenientes da extração do DNA plasmidial foram utilizadas nos passos seguintes da clonagem.

O vetor de expressão pET28a e os plasmídeos foram digeridos utilizando as enzimas de restrição *NdeI* e *XhoI* (ThermoScientific). As reações de dupla digestão para cada plasmídeos foram montadas de acordo com a Tabela 1 e incubados a 37 °C por 4 horas. O vetor de expressão pET28a foi linearizado com as mesmas enzimas de restrição seguindo o

protocolo da Tabela 1, porém foi acrescentado 1 U de fosfatase alcalina FastAP (ThermoFischer Scientific) à reação para impedir a recircularização do vetor. Após o tempo de incubação as reações foram aplicadas em gel de agarose e as bandas foram cortadas do gel com o auxílio de uma lâmina de bisturi e posteriormente purificados utilizando o GeneJET Gel Extraction and DNA Cleanup Micro Kit (ThermoFischer Scientific) seguindo o protocolo estabelecido pelo fabricante.

Reagente	Volume
Tampão R (10x)	5 µL
BSA	1 µL
NdeI*	25 U
XhoI*	25 U
DNA (3ug)	-
Água	-
Volume final	50 µL

 Tabela 1- Reação de digestão dupla utilizando as enzimas de restrição NdeI e XhoI

* Enzimas de restrição utilizadas da Fabricante ThermoFischer Scientific

A inserção dos genes no vetor pET28a linearizado foi realizada utilizando T4 DNA ligase (Invitrogen) seguindo o protocolo descrito na Tabela 2. Para calcular a quantidade de vetor e inserto necessário foi utilizado a equação ⁽¹⁾ usando uma proporção de vetor para inserto de 1:3. As reações foram incubadas a temperatura ambiente por 18 horas. Em seguida 5 μ L da reação foram utilizados para transformação por choque térmico em 50 μ L de células *E. coli* DH5 α , como descrito acima, e incubados a 37 °C por 18 horas. Uma colônia de cada construção foi inoculada em 10 mL de meio LB contendo 50 μ g/mL do antibiótico seletivo canamicina e esta foi crescida a 37 °C por 18 horas. Em sequência foi realizada a extração dos plasmídeos e uma amostra de cada construção foi enviada a *facility* no Instituto de Química-USP para serem sequenciadas. Após confirmação dos resultados do sequenciamento, foi realizado teste de expressão heteróloga das proteínas alvos.

$$\frac{ng \, (vetor) \times Kb \, (inserto)}{Kb \, (vetor)} = razão \, molar \, \frac{inserto}{vetor} \tag{1}$$
Reagente	Volume
Tampão de ligação (5X)	4 μL
Vetor	-
Inserto	-
T4 DNA ligase	1 U
ATP	1 µL
Água	-
Volume final	20 µL

Tabela 2- Reação de ligação do inserto no vetor de expressão.

3.2 Eletroforese em Gel de Agarose

Todas as amostras de DNA foram analisadas em gel de agarose 1% ou 0,8%, se uma etapa subsequente de purificação da amostra contida no gel fosse necessária. Os géis foram preparados com TAE (40 mM Tris-base pH7,6, 20 mM ácido acético e 1 mM EDTA) e utilizou-se o fluoróforo EnviroSafe (Helixx) 0,05 μL/mL para visualização posterior das bandas. Para a preparação das amostras, foi utilizado o tampão para amostra de DNA 6X (30% v/v glicerol, 0,25% w/v azul de bromofenol e 0,25% w/v xileno cianol FF). A cuba do aparelho de eletroforese horizontal MUPID-exU (Mupid) foi preenchida com tampão TAE e as amostras e marcador molecular GeneRuler 1Kb (ThermoFischer Scientific) ou DNA Ladder 1Kb (NEB) foram aplicadas no gel. A corrida eletroforética foi realizada a 70 V por 50 minutos. Após o término de cada corrida, utilizou-se o transiluminador ChemiDoc-It (UVP) com luz UV para visualização e fotodocumentação dos géis.

3.3 Teste de Expressão em Pequena Escala

Para determinar as melhores condições de expressão foram realizados testes de expressão com as proteínas alvo no intuito de obter-se o melhor rendimento proteico. As diferentes condições testadas estão descritas na Tabela 5. Após transformação dos plasmídeos em cada cepa, também foram realizadas as transformações dos plasmídeos vazios como controle negativo. Após incubação de 16 horas das placas de petri a 37 °C foram obtidas colônias. Uma colônia de cada placa foi inoculada em um *erlen-meyer* contendo 20 mL de meio LB suplementado com o antibiótico apropriado de cada cepa/ vetor e incubada a 37 °C sob agitação constante de 200 rpm por 4 horas. Após completar esse tempo foi retirado 1 mL de cada cultura que foi centrifugado a 14.000 *g* por 1 minuto, o sobrenadante foi descartado e

o *pellet* foi guardado a -20 °C. O restante de cada cultura foi repartido em 3 tubos com 6 mL cada. Em cada tubo foi adicionado isopropil-β-D-tiogalactosídeo (IPTG) de tal forma que para cada cepa foram testados 3 concentrações finais de IPTG. Essas culturas foram repartidas novamente em 2 tubos contendo 3 mL da cultura e cada tubo foi acondicionado a duas temperaturas, sendo que um foi mantido a 37 °C por 4 horas e outro, a 18 °C por 18 horas. Em seguida, foram retirados 1 mL de cada cultura, que foi centrifugado a 14.000 *g* e o sobrenadante foi descartado. Cada pellet foi ressuspendido com 100 µL de BugBuster (Novagen) e 0,5 µL de uma mistura de lisozima (10mg/mL) e DNAse (1 mg/mL) e mantidos em temperatura ambiente por 30 minutos. Logo em seguida o lisado foi centrifugado a 14.000 *g* por 3 minutos. O sobrenadante, equivalente à fração solúvel, foi colocado em outro tubo enquanto que no *pellet* foram adicionados 50 µL de BugBuster, que foi centrifugado por SDS-PAGE.

Tabela 3- Condições testadas no teste de expressão.

Cepas	Temperaturas	IPTG	Tempo*
BL21(DE3)	37 °C	0,2 mM	4 horas
BL21(DE3)Pgro7	18 °C	0,4 mM	18 horas
Tuner(DE3)		0,8 mM	

* Tempo total da expressão após indução com IPTG

3.4 Expressão Heteróloga de Proteínas Recombinantes

A partir dos resultados dos testes de expressão foram escolhidas as melhores condições para serem reproduzidas em grande escala. Inicialmente foi preparado um préinóculo no qual uma colônia foi inoculada em meio LB contendo o antibiótico seletivo apropriado. As culturas foram incubadas por 16 horas a 37 °C sob agitação constante de 200 rpm. Essa pré-cultura foi utilizada como inóculo (10 mL/L) em meio LB contendo os respectivos antibióticos e incubada a 37 °C sob agitação de 200 rpm até a densidade ótica (A₆₀₀) atingir um valor entre 0,6 e 0,8. Em seguida, a temperatura foi alterada para 18 °C. Após 30 minutos foi adicionado isopropil- β -D-tiogalactosídeo (IPTG) e a cultura foi mantida a 18 °C, 200 rpm, por 18 horas. Posteriormente, as culturas foram centrifugadas a 12.000 *g* por 30 min, o sobrenadante foi descartado e o *pellet* armazenado a -80 °C. A densidade celular foi medida a 600 nm com o espectrofotômetro óptico DiluPhotometer (IMPLEN).

3.5 Expressão de Proteínas Recombinantes em Biorreator

Foi realizada a expressão heteróloga de proteínas em biorreatores em colaboração com a Dra. Marilda Keico Taciro e o aluno Kaíque Josuel do Laboratório de Bioprodutos, no Departamento de Microbiologia - ICB II. Essa expressão foi preparada e executada seguindo o protocolo de Bernard e Payton (2001) e as condições, como a cepa e concentrações de IPTG, foram escolhidas de acordo com o resultado dos testes de expressão. Uma pré-cultura de 400 mL de meio LB foi preparada para as construções pET28a-TylM2 transformada em E. coli BL21(DE3)pGRO7 e pET28a-TylM3 transformada em Tuner(DE3), nas quais foram adicionados os antibióticos canamicina/cloranfenicol e canamicina, respectivamente. Essas pré-culturas foram incubadas a 37 °C sob agitação constante de 200 rpm por 16 horas. Em seguida, as pré-culturas foram transferidas para biorreatores, modelo EZ da Applikon com 5L de capacidade de volume total da dorna, contendo 4L de meio ECPM1 (Tabela 4). Oxigênio dissolvido no reator foi mantido acima de 20% da saturação através da variação da frequência de agitação (400-1100 rpm). A vazão de ar foi mantida em 0,5 vvm nas primeiras 5 horas de ensaio e posteriormente mantida em 0,75 vvm, o pH controlado em 7,0 com adição de ácido fosfórico (85% v/v) e hidróxido de amônia (28% v/v) e temperatura controlada a 37 °C. No experimento com a linhagem BL21(DE3)pGRO7 foi adicionado L-arabinose na concentração final de 0,5 mg/ml com 1,5 hora de experimento. Em ambos os ensaios, após 5 horas de cultivo, quando a densidade ótica (A₆₀₀) atingiu 4,0 foi adicionado IPTG a uma concentração final de 0,2 mM e a temperatura alterada para 18 °C e mantido assim por mais 16 horas. Por fim, toda cultura do biorreator foi centrifugada a 12.000 g por 30 minutos, o sobrenadante foi descartado e o *pellet* fracionado em tubos separados, foram armazenados a -80 °C.

Meio ECPM1	
Reagentes e soluções	Quantidade
K ₂ HPO ₄	4 g
KH_2PO_4	1 g
NH_4Cl	1 g
K_2SO_4	2,4 g
CaCl.2H ₂ O	132 mg
Ácidos casamino	20 g

Tabela 4- Composição do meio ECPM1 e solução traços. Fonte: Bernard; Payton (2001)

Extrato de levedura	3 g
Glicerol	40mL
Solução traço	10 mL
Água	-
Volume final	800 mL

Composição da s	solução traço
Reagentes	Quantidade
EDTA	5 g
FeCl ₃ .6H ₂ O	0,5 g
ZnO	0,05 g
CuCl ₂ .2H ₂ O	0,01 g
Co(NO ₃) ₂ .6H ₂ O	0,01 g
(NH4)6Mo7O ₂₄ .4H ₂ O	0,01 g

3.6 Extração e Purificação das Proteínas

Os pellets congelados foram resuspendidos e incubados por 30 minutos no gelo com tampão lise (50 mM Tris-HCl pH 7,8, 200 mM NaCl, 10% v/v glicerol, 0,1% v/v Thesit e 5 mM β-mercaptoetanol) com 0,5 mM PMSF, 100 µg/mL de lisozima e 3U DNAse. As células foram rompidas por sonicação (BransonSonifier) em amplitude de 70%, com pulsos de 15 segundos e 30 segundos de descanso por 15 minutos. Para a purificação das proteínas foi utilizado o FPLC Äkta Purifier (GE Healthcare Life Sciences). O lisado foi centrifugado a 35.000 g por 1 hora e o sobrenadante foi injetado em uma coluna HisTrap IMAC HP 5 mL (GE) pré empacotada e carregada com íons de níquel. Essa coluna foi lavada e equilibrada com 10 volumes de coluna (10 cv) do tampão de ligação (50 mM Tris-HCl pH 7,8, 200 mM NaCl, 10% v/v glicerol, 0,1% v/v Thesit e 5 mM β-mercaptoetanol). As proteínas ligadas à coluna foram eluídas a um gradiente crescente linear de imidazol de 5 a 100% com o tampão de eluição (50 mM Tris-HCl pH 7,8, 200 mM NaCl, 10% v/v glicerol, 0,1% v/v Thesit, 5 mM β-mercaptoetanol e 500 mM imidazol). Após a purificação as frações referentes à eluição da proteína foram analisadas em gel SDS-PAGE 10%. Os picos contendo a proteína alvo foram concentrados em tubos concentradores AmiconUltra-15 (EMD Millipore). Em seguida foi realizada uma cromatografia por exclusão de tamanho. Para a aplicação da amostra na coluna a proteína foi concentrada a volume final de 2 mL e injetada na coluna HiLoad 16/600 Superdex 200 (GE Healthcare Life Sciences). O resultado da cromatografia de exclusão por tamanho foi verificado em gel SDS-PAGE 10% e as amostras contendo a proteína pura, foram concentradas em tubos Amicon Ultra-4 (EMD Millipore) até atingir a concentração mínima de 10 mg/mL. A concentração da proteína foi medida a 280 nm utilizando o equipamento Biochrom Libra S50.

3.7 Eletroforese de Proteínas

Amostras proteicas foram analisadas por eletroforese em gel de poliacrilamida contendo dodecil sulfato sódio (SDS) e em gel nativo.

<u>3.7.1 SDS-PAGE</u>

Amostras provenientes das cromatografias por afinidade a íons metálicos imobilizados e a cromatografia por exclusão de tamanho foram analisadas sob condições desnaturantes em géis de poliacrilamida contendo SDS. Os géis foram preparados conforme protocolo de Sambrook e Russel (2000). A composição dos géis de empilhamento e de resolução está descrita na tabela 5. As amostras de proteína foram preparadas adicionando-se tampão de amostra 4X (Glicerol 40% v/v, 250 mM Tris-HCl pH 6.8, 8% w/v SDS, 00,4% de azul de bromofenol e 5% β-mercaptoetanol) e em seguida foram aquecidas a 98 °C por 5 minutos. Logo após, as amostras foram aplicadas no gel e foi utilizado o marcador de peso molecular Precision Plus Dual Color 250 kDa (BioRad). Utilizou-se o aparelho de eletroforese vertical Mini-PROTEAN Tetra Cell (Biorad) cuja cuba foi preenchida com tampão de corrida (250 mM Tris-HCl, 1,92 M Glicina e 1% w/v SDS). A eletroforese foi iniciada com 70 V por 15 minutos e em seguida a voltagem foi aumentada para 170 V. Após 1 hora os géis foram corados com corante Coomassie Brilliant Blue G-250 e descorados conforme protocolo de Lawrence e Besir (2009).

Tabela 5- Composição de géis de empacotamento e resolução para SDS-PAG	GE
--	----

Reagente	Gel de empilhamento 5%	Gel de resolução 10 %
Bis-acrilamida 30%	830 μL	3,3 mL
1,5 M Tris-HCl pH 8,8	-	2,5 mL
1 M Tris-HCl pH 6,8	630 µL	-
Água	3,4 mL	4 mL
SDS 10 %	50 µL	100 µL
*APS 10%	50 µL	100 µL

**TEMED	5 µL	4 µL
Volume final	5 mL	10 mL

* Persulfato de amônia (APS) e ** N,N,N',N'-tetramethylenediamine (TEMED) são adicionados somente no final

<u>3.7.2 BN-PAGE</u>

Para análise de complexos de proteínas sob condições nativas foi preparado gel nativo azul (BN-PAGE) com auxílio do Dr. Carlos Gustavo Baptista e do Prof. Ariel Silber do Departamento de Parasitologia do Instituto de Ciências Biomédicas da Universidade de São Paulo. Esse método é utilizado para determinar a massa de proteínas de baixo a alto peso molecular (10 MDa), determinar o estado oligomérico e identificar interações proteínaproteína (WITTIG; BRAUN; SCHÄGGER, 2006). Utilizou-se um gradiente linear (5% a 15%) de acrilamida. Os tampões de gel, catodo e anodo, juntamente com a composição do gel estão descritos na tabela 6 e 7 abaixo.

Tabela 6- Composição dos géis de empacotamento e de resolução para BN-PAGE

Reagentes	Gel de empilhamento 3,5%	Gel de resolução 5%	Gel de resolução 15%
Bis-acrilamida 30%	0,25 mL	0,5 mL	1,04 mL
Tampão de gel 3X	1 mL	1,66 mL	1,11 mL
Água	1,75 mL	2,8 mL	0,6 mL
Glicerol 100%	-	-	0,56 mL
*APS 10%	25 μL	40 µL	24 µL
**TEMED	3 µL	4 μL	3 µL

* Persulfato de amônia (APS) e ** N,N,N',N'-tetramethylenediamine (TEMED) são adicionados somente no final

Labera /				
	Reagente	Tampão de gel 3X	Tampão anodo	Tampão catodo
	Bis-Tris pH 7,0	150 mM	50 mM	1,5 mM
	Ácido aminocapróico	1,5 M	-	-
	Tricina	-	50 mM	-
	Coomassie Blue G-250	-	-	0,02 %

Tabela 7- Composição dos tampões para BN-PAGE

Os tampões A (0,75 M ácido aminocapróico e 50 mM Bis-Tris pH 7,0) e B (5% w/v Coomassie Brilliant Blue G-250 e 0,75 M ácido aminocapróico) foram utilizados na preparação de amostras, na qual adicionou-se 5 μ L de cada tampão à 15 μ L de proteína (50 a 100 μ g), totalizando 20 μ L de amostra por poço. Utilizou-se 5 μ L do marcador de peso molecular NativeMark Unstained Protein Standard (Invitrogen). As amostras foram incubadas em gelo até o momento de montar o gel no aparelho de eletroforese vertical Mini-PROTEAN Tetra Cell (Biorad). A cuba foi preenchida com tampão anodo e a parte interna onde o gel é mantido foi preenchida com tampão catodo. A eletroforese foi iniciada a 80 V por 30 minutos e posteriormente a voltagem foi aumentada para 300 V. Após aproximadamente 3-4 horas, o gel foi corado e descorado como descrito acima para SDS-PAGE.

3.8 Dicroísmo Circular

Para as medidas de espectroscopia de dicroísmo circular foi utilizado o espectropolarímetro J-810 (Jasco), equipado com um Peltier Type Control System PFD425S, que controla a temperatura, para realizar uma caracterização da estrutura secundária das proteínas, no Centro de Biotecnologia, Instituto Butantã- São Paulo. Utilizou-se 120 μ L das amostras numa concentração de 1 mg/mL. Antes das medidas serem efetuadas, as amostras foram centrifugadas a 15.000 *g* durante 5 minutos a 4 °C para remoção de qualquer partícula de agregados. Após isso, as amostras foram pipetadas cuidadosamente em cubetas de quartzo com caminho óptico de 0,1 cm. Para a coleta dos espectros, foi utilizado um comprimento de onda de 190 a 260 nm, para cada medida foram feitas 20 acumulações a uma temperatura constante de 10 °C com intervalo entre as coletas de 0,5 nm. Antes de realizar as aquisições das amostras, foram feitas aquisições dos tampões de cada amostra para posteriormente ser utilizado para tratar os dados.

3.9 Espalhamento Dinâmico de Luz

Para o ensaio de espalhamento de luz dinâmica, foi utilizado o equipamento DynaPro NanoStar (Wyatt Technology) presente no Laboratório de Proteínas e Biologia Redox do Prof. Dr. Luis Netto no Instituto de Biociências– USP. As amostras foram centrifugadas a 15 000 g durante 5 minutos a 4 °C para remoção de impurezas ou partículas insolúveis em suspenção. Para a medição de cada amostra (1 mg/mL) foi utilizado uma cubeta de quartzo com caminho de 10 mm, a uma temperatura de 25 °C.

3.10 Cristalização De Proteínas

Para cristalização das proteínas, foi utilizado o sistema de gota sentada, com os *screens* padronizados e comercializados pelas empresas Hampton Research e Jena Bioscience. Inicialmente, 30 μ L de cada condição dos kits selecionados foram pipetados em placas de 96 poços MRC2 (swissci). Para a distribuição da proteína nas placas foi utilizado o robô Oryx4 (Douglas Instruments), que montou as gotas sentadas com um volume final de 0,60 μ L, sendo 0,30 μ L de proteína concentrada a 10 mg/mL e 0,30 μ L da solução do reservatório. Após a obtenção dos *hits*, as condições escolhidas foram otimizadas variando parâmetros como pH e concentração de proteína ou do agente precipitante. Para isso, foram pipetados 300 μ L de cada condição otimizada em placas VDX de 24 poços (Hampton Research). Em lamínulas siliconizadas foram montadas 3 gotas, variando a proporção entre proteína e a condição do reservatório (1:1; 1:2 e 2:1 respectivamente). O sistema foi vedado utilizando óleo de imersão de alta viscosidade nas bordas de cada poço. E essas lamínulas contendo as 3 gotas foram invertidas sobre cada poço de forma que as gotas ficassem penduradas. Todas as placas foram mantidas a 18 °C e posteriormente, observadas diariamente em uma lupa estereoscópica para verificar o crescimento de cristais.

3.11 Coleta de Dados de Difração de Raio-X

As coletas de dados de difração de raios-X foram realizadas na linha MX2 do Laboratório Nacional de Luz Síncrotron (LNLS) em Campinas- São Paulo e BL9-2 do Stanford Synchrotron Radiation Lightsource (SSRL) em Stanford- Estados Unidos, ambas dedicadas a cristalografia de proteínas. Todos os cristais foram crioprotegidos com uma solução de 30% v/v glicerol preparada a partir da condição de cristalização. Em seguida os cristais foram pescados com *loop* de *nylon* e transferidos para a solução crioprotetora por 15 segundos e em seguida montado no goniômetro onde os cristais foram mantidos em vapor de nitrogênio líquido na temperatura de -173 °C. O detector foi posicionado a 100 mm do cristal e a qualidade da difração é avaliada pela aquisição de uma imagem a 0 e 90° de oscilação. Após isso, foram coletadas entre 1800-7200 imagens no qual o cristal foi exposto ao feixe de raios-X por um tempo total de 0,1-1 segundo, rotacionando entre 0,05-0,2°.

3.12 Determinação da Estrutura Cristalográfica

Os dados de difração de raios-X foram processados utilizando o programa XDS (KABSCH, 2010), que localiza as reflexões do padrão de difração e fornece algumas

estimativas como parâmetros de célula unitária e grupo espacial a qual o cristal pertence. Para estimar o número de moléculas presente na unidade assimétrica do cristal foi utilizado o programa Matthews_coef. O escalonamento dos dados foi realizado pelo programa Aimless (EVANS; MURSHUDOV, 2013) e a tentativa da determinação da estrutura foi realizada nos programas Phaser (McCOY; et al., 2007) e MOLREP (VAGIN; TEPLYAKOV, 1997), ambos presentes no pacote de programas CCP4 (BAILEY, 1994).

3.13 Espalhamento de Raios-X a Baixo Ângulo

A coleta de dados de TylM2 (1 mg/mL) em tampão 50 mM Tris-HCl, 150 mM NaCl, 10% v/v glicerol, 0,1 % v/v Thesit e 5 mM β -Mercaptoetanol, pH 7,8 a partir do espalhamento de raios-X de amostras não cristalinas foi realizado na linha SAXS1 no Laboratório Nacional de Luz Síncrotron (LNLS) situado em Campinas, São Paulo. As aquisições foram feitas utilizando o detector pilatus 300K (84 mm x 107 mm), com comprimento de onda $\lambda = 1,55$ Å e um porta amostras com 1 mm de espessura de janelas de mica em temperatura ambiente. As amostras foram centrifugadas a 13 000 g por 3 minutos a 4 °C antes de iniciar a coleta. Foram obtidas curvas de espalhamento a partir da mica vazia, com tempo de exposição de 10 segundos; da mica com o tampão de cada amostra de proteína a um tempo de exposição de 300 segundos e por fim, da mica com a solução de proteínas com tempos de exposição de 30 segundos.

A integração dos dados, normalização da intensidade do feixe e a subtração do espalhamento do tampão foi realizado pelo programa disponibilizado na linha. Para a análise de dados foi utilizado o pacote de programa ATSAS (disponível em: http://www.embl-hamburg.de/biosaxs/software.html). O programa Autorg (PETOUKHOV et al., 2007) foi utilizado para uma avaliação inicial da qualidade dos dados, e para obter informações iniciais para o raio de giro (R_g) e a intensidade de espalhamento a ângulo zero (I/(0)). Esses mesmos valores foram comparados no programa Primus (KONAREV et al., 2003) onde foram geradas as curvas de Guinier e Krakty.

3.14 Modelagem Molecular por Homologia

Foi utilizado o programa MODELLER v9.17 (WEBB; SALI, 2014) para a construção dos modelos das proteínas TylM2 e TylM3. A busca por homólogos foi realizada com a ferramenta BLAST (hospedada no National Center Biotechnology Information – NCBI,

disponível em https://www.ncbi.nlm.nih.gov/) contra o banco de dados de proteínas (PDB, disponível em http://www.rcsb.org/pdb/home/home.do), utilizando as sequências de resíduos de aminoácidos referente às proteínas de interesse obtidas no European Nucleotide Archive (ENA, disponível em http//:www.ebi.ac.uk/ena). A ferramenta BLAST permite realizar buscas de sequências de nucleotídeos (BLASTN) ou de proteínas (BLASTP) contra um banco de dados específico para encontrar similaridades locais através do alinhamento entre sequências (ALTSCHUL et al., 1990). Foram escolhidas proteínas que apresentassem homologia sequencial acima de 30% com as proteínas de interesse. Após construção dos modelos, seus níveis de qualidade e confiabilidade foram verificados utilizando os programas ProSA-web (SIPPL, 1993; WIEDERSTEIN; SIPPL, 2007), PROCHECK (LASKOWSKI et al., 1993) e VERIFY3D (BOWIE; LÜTHY; EISENBERG, 1990; LÜTHY; BOWIE; EISENBERG, 1992) e para visualização dos modelos foi utilizado o programa Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

4 RESULTADOS E DISCUSSÃO

4.1 Clonagem, Expressão e Purificação

4.1.1 Sub-clonagem, Expressão e Purificação das Proteínas TylM2 e TylM3

Os genes sintéticos (GenScript) para a glicosiltransferase *tylm2* e sua proteína auxiliar *tylm3* foram obtidos clonados em pUC57. Esses genes foram subclonados em vetor de expressão pET28a e transformados em *E. coli* DH5 α . *Escherichia coli* é uma bactéria gram negativa escolhida como modelo para trabalhar com DNA recombinante devido à sua capacidade de crescimento rápido, apresentando um tempo de geração de 20 a 30 minutos. Quase todas as cepas utilizadas em estudos com DNA recombinante são derivadas da *E. coli* K-12, A cepa DH5 α é amplamente utilizada para clonagem devido à algumas características em seu genótipo, dentre elas destacam-se duas: a capacidade de manter a estabilidade do inserto dentro da célula hospedeira que se deve à mutação do gene *recA*, que impede a recombinação do DNA recombinante (inserto) com o DNA da célula hospedeira; e a capacidade de aumentar o rendimento e qualidade do plasmídeo inserido através da mutação *endA1*, que impede a ação de endonucleases não específicas de reconhecerem e degradarem DNA exógeno (CASALI, 2003).

Para confirmação da clonagem foi realizada uma digestão do plasmídeo (Figura 9) onde foi possível identificar duas bandas entre 1000 e 1500 pb com tamanho relacionado aos genes *tylm2* (1368 pb) e *tylm3* (1281 pb). Uma amostra de cada construção foi enviada para serem sequenciadas no *facility* do IQ-USP, confirmando a presença dos genes nos plasmídeos.



Figura 9- Digestão dos plasmídeos pET28a-*tylM2* e pET28a-*tylM3* M- GeneRulerTM 1Kb DNA Ladder (ThermoFischer Scientific). 1- *tylm2* (1368 pb) e 2- *tylm3* (1368 pb). Cetas brancas indicam posicionamento das bandas referente a cada gene no gel de agarose 1%.

Com a confirmação da subclonagem foram iniciados os testes de expressão das proteínas de interesse em pequena escala. Condições, tais como temperatura e concentração do indutor representam parâmetros que afetam a quantidade de proteína recombinante produzida, bem como sua solubilidade. A partir das análises de SDS-PAGE foi possível determinar que a melhor condição para a TylM2 (Figura 10a) foi obtida na cepa de E. coli BL21(DE3) transformada com o plasmídeo de chaperonas pGro7 (Takara), indução de 0,2 mM de IPTG com temperatura de cultivo de 37 °C posteriormente baixado a 18 °C por 16 horas. O plasmídeo pGro7 codifica as chaperonas de E. coli GroEL (60 Kda) e GroES (10 Kda), permitindo a co-expressão com a proteína de interesse. Esse sistema de chaperonas forma um complexo que se assemelha a um barril vazio, formado por dois anéis heptaméricos de GroEL um sobre o outro e a tampa, formada por um heptâmero de GroES. As proteínas se ligam temporariamente ao barril formado por GroEL e a hidrólise de ATP fornece energia para os ciclos de enovelamento das proteínas parcialmente enoveladas, assim auxiliando no seu enovelamento correto e consequentemente mantendo-as solúveis (MARQUES, 2012). Para a produção de proteínas recombinantes sob controle do promotor T7, são utilizadas cepas contendo o lisógeno de fago λ DE3 que codifica a RNA polimerase T7.

A produção da proteína auxiliar TylM3 se provou mais difícil, pois praticamente em todas as condições testadas a proteína encontrava-se na fração insolúvel. A condição em que foi obtido um quantidade insignificativamente maior em fração solúvel foi com a cepa Tuner(DE3), indução de 0,2 mM de IPTG com temperatura de cultivo de 37 °C posteriormente baixado a 18 °C por 16 horas (Figura 10b). Essa cepa, derivada de BL21 possui uma mutação *lacY1*, o que permite uma entrada uniforme de IPTG na célula. Ao utilizar níveis mais baixos de IPTG na expressão é possível aumentar a solubilidade de proteínas (KHLEBNIKOV; KEASLING, 2002).



Figura 10- Análise do teste de expressão para as proteínas TylM2 e TylM3. Proteína TylM2 (50 kDa) transformada na cepa BL21(DE3)pGRO7 crescida a 37 °C por 4 horas (a) e em 18 °C por 16 horas (b). E análise da expressão da proteína TylM3 (46 kDa) transformada na cepa Tuner(DE3) crescida a 37 °C por 4 horas (c) e em 18 °C por 16 horas (d). M- Marcador de peso molecular Precision Plus Duo Color (BioRad) 250 kDa, 1-não induzido; 2-fração insolúvel com 0,2 mM de IPTG; 3-fração solúvel com 0,2mM de IPTG; 4-fração insolúvel com 0,4 mM de IPTG; 5-fração solúvel com 0,4mM de IPTG; 6-fração insolúvel com 0,8 mM de IPTG; 7-fração solúvel com 0,8mM de IPTG. Setas indicam presença das proteínas de acordo com seus pesos moleculares.

Quando comparamos a purificação da proteína TylM2 produzida em *shaker* com a produzida no biorreator foi possível verificar que não houve aumento no rendimento em relação à proteína de interesse, no entanto houve uma produção de outras proteínas da *E. coli* de forma acentuada, o que comprometeu a pureza da amostra. As vantagens de se utilizar um biorreator para expressão de proteínas recombinantes é o controle que se tem sob condições que permitem atingir densidades altas de células tais como o nível de oxigenação, pH e nutrição (SIVASHANMUGAM et al., 2009). Em ambos os ensaios em biorreator, as proteínas exibiram uma densidade celular maior ao final da expressão quando comparado com a de *shaker*, apresentando peso úmido de aproximadamente 45 g/L e 3,5 g/L, respectivamente. Sendo assim um aumento da densidade celular de aproximadamente 7 vezes. Na cromatografia por afinidade a íons metálicos imobilizados, a TylM2 eluiu com aproximadamente 150 mM de imidazol em ambas as purificações e análises em SDS-PAGE constataram a diferença entre a proteína produzida em *shaker* (Figura 11a) e em biorreator (Figura 11b), evidenciando uma maior quantidade de contaminantes produzida durante o cultivo em biorreator.



Figura 11- Análise da proteína TylM2 (50 kDa) produzida em *shaker* (a) e em bioreator (b): A esquerda, os cromatogramas das purificações, seta indica eluição da proteína a aproximadamente 30% de imidazol, equivalente a 150 mM. A direita, SDS-PAGE 10% após cada purificação. Setas indicam presença da proteína em fração solúvel. M-Marcador de peso molecular Precision Plus Duo Color (BioRad) 250 kDa, 1-*pellet*; 2-lisado; 3-*flow through*; 4-5% imidazol; 5-14- eluição da proteína.

Após a produção de TylM3 em *shaker*, foi realizada a cromatografia por afinidade. Como pode ser visto na Figura 12a a proteína eluiu com aproximadamente 100 mM de imidazol e análises de SDS-PAGE revelaram que, mesmo tendo uma quantidade pequena na fração solúvel, não seria o suficiente para etapas subsequentes de cristalização, pois a amostra ainda precisaria passar por outra técnica de cromatografia para aumentar sua pureza e com cada processo a mais, há perda da quantidade de amostra. Foi então realizado a expressão da proteína em biorreator para obter uma maior densidade celular e com isso, obter maior quantidade da proteína após sua extração e purificação. Porém, ao final da expressão as medidas de oxigênio dissolvido tinham aumentado. A indução da produção de proteína recombinante é realizada antes que a cultura alcance a fase estacionária de crescimento. Nesta fase, não há um aumento no número de células e a quantidade de oxigênio dissolvido se mantém estável, pois as células presentes na cultura estão metabolicamente ativas. Deste modo, se houvesse um aumento na taxa de oxigênio dissolvido, indicaria que em algum momento haveria uma diminuição da população devido a morte celular. Mesmo assim, essa cultura foi centrifugada, e as células rompidas para liberação da proteína da fração citoplasmática. Após a purificação, foi possível verificar que a maior parte da proteína estava presente na fração insolúvel, impossibilitando seu uso para a etapa seguinte de cristalização (Figura 12b). Dados obtidos das amostras que foram coletadas durante o ensaio para quantificar (A₆₀₀) a densidade celular de cada cultura foram plotadas e pode ser visto na Figura 13. Após aproximadamente 9 horas de cultivo da linhagem Tuner(DE3) contendo a construção pET28a-*tylM3* houve uma diminuição da densidade celular e por isso foi detectado um aumento da taxa de oxigênio dissolvido no biorreator.



Figura 12- Análise da proteína TylM3 (46 kDa) produzida em *shaker* (a) e em bioreator (b). À esquerda, os cromatogramas das purificações por afinidade, seta indica eluição da proteína a aproximadamente 20% de imidazol, equivalente a 100 mM. À direita, SDS-PAGE 10% após cada purificação. Setas indicam presença da proteína em fração solúvel. M- Marcador de peso molecular Precision Plus Duo Color (BioRad) 250 kDa, 1-*pellet*; 2-lisado; 3-*flow through*; 4- 5% imidazol; 5-14- eluição da proteína.



Figura 13- Curva da densidade celular das cepas BL21(DE3)pGRO7 e Tuner(DE3), utilizados para expressão das proteínas TylM2 e TylM3, respectivamente. A seta indica o momento no qual foi adicionado 0,2 mM de IPTG aos cultivos.

A expressão de proteínas recombinantes em larga escala pode resultar no enovelamento incorreto de proteínas e sua consequente degradação por proteases celulares. As proteínas parcialmente enoveladas podem ainda associarem entre si formando um agregado de proteínas biologicamente inativo denominado corpo de inclusão. É possível trabalhar com proteínas que estão em corpo de inclusão através de *refolding*, porém essa metodologia não é muito viável devido ao baixo rendimento e a incerteza de que a proteínas assumirá o correto enovelamento. Outra tentativa de aumentar a solubilidade de proteínas recombinantes é diminuir a concentração do indutor e diminuir a temperatura da cultura, deixando-a por mais tempo. Mesmo os genes codificando as proteínas TylM2 e TylM3 terem sido otimizados para expressão em *E. coli*, o rendimento final de ambas se mostrou baixo, mesmo após a utilização das estratégias citadas acima.

Estudos com homólogas da TylM2 e TylM3, DesVII e DesVIII, respectivamente foram realizadas por Borisova e Liu, (2010), nos quais foram relatados a dificuldade na produção da proteína auxiliar DesVIII solúvel, entretanto ao utilizar um vetor codificando a proteína fusionada MBP a solubilidade da proteína aumentou, assim como seu rendimento. Um problema semelhante de baixa expressão e rendimento foi relatado para a glicosiltransferase EryCIII por Lee et al. (2004). Porém esse problema foi solucionado com a utilização de chaperonas GroEL/GroES. Já, Moncrieffe et al. (2012) relataram que a co-expressão da EryCIII com GroEL e GroES de *E. coli* produziu uma quantidade relativamente baixa de proteína solúvel e após a purificação da glicosiltansferase, a pureza da amostra foi

comprometida pela contaminação de GroEL, porém ao utilizarem as chaperonas de *Streptomyces coelicolor* GroEL1, GroES e GroES2 foi obtida a proteína solúvel sem contaminante por chaperonas. Esse problema de contaminação foi também relatado por Lee et al. (2004), sugerindo que a presença de chaperonas após purificação da proteína seja indicativo de que há proteínas não corretamente enoveladas.

4.2 Caracterização Estrutural da Proteína TylM2

4.2.1 Gel Nativo

Na eletroforese convencional em condições desnaturantes SDS-PAGE as proteínas são separadas de acordo com a massa molecular relativa. Já na eletroforese em condições não desnaturantes, as proteínas possuem uma carga líquida distinta e são separadas de acordo com suas propriedades eletroforéticas no gel. Ou seja, se há mais de um estado oligomérico na amostra, cada oligômero terá uma carga distinta e assim, percorre o gel de forma diferente.

O perfil da cromatografia por exclusão de tamanho é composto por um único pico com volume de retenção de 60-70 mL. Esse volume é consistente com massas moleculares de aproximadamente 200 KDa. O peso molecular da proteína TylM2 em estado monomérico é de aproximadamente 50 KDa, então para corroborar os resultados da cromatografia por exclusão de tamanho, foi realizado um gel nativo para analisar a massa molecular da proteína em condições não-desnaturantes (Figura 15). Foi possível verificar que a proteína TylM2 está presente em dois estados oligoméricos, há uma banda fraca próximo à 66 KDa, referente ao provável estado monomérico (50 KDa), e outra abaixo de 242 KDa, possivelmente referente a um tetrâmero.



Figura 14- BN-PAGE 5-15% da proteína TylM2 (50 KDa). M- NativeMark Unstained Protein Standard 1236 KDa (life technologies); 1-4 amostras da proteína TylM2 purificada.

4.2.2 Dicroismo circular

No espectropolarímetro, a luz circularmente polarizada é produzida quando a luz proveniente de uma fonte passa por dois prismas e uma série de espelhos e fendas até ser focalizada por uma lente, passando por um filtro e chegando ao modulador, onde a luz linearmente polarizada é desviada a direita e a esquerda, produzindo assim luz circularmente polarizada. Em seguida, essa luz passa por um obturador e pelo porta amostra até finalmente ser detectado pelo fotomultiplicador, um equipamento que converte o sinal de luz em corrente elétrica. Dessa forma, quando uma molécula contém um ou mais grupos funcionais que absorvem luz, há uma diferença entre a absorção de luz circularmente polarizada à direita e à esquerda. Essa diferença na absorção gera o sinal de dicroísmo circular em um determinado comprimento de onda.

Foi obtido o espectro para a proteína TylM2 em 50 mM Tris-HCl, 150 mM NaCl, 10% v/v glicerol, 0,1% v/v Thesit e 5 mM β -Mercaptoetanol, pH 7,8. A partir do gráfico (Figura 16) foi possível confirmar que a amostra proteica está enovelada. Além disso, os dados gerados pelo experimento foram inseridos no servidor K2D3 (disponível em: http://cbdm-01.zdv.uni-mainz.de/~andrade/cgi-bin/k2d3/k2d3_set1.pl) onde foi possível obter informações sobre a composição de estruturas secundárias. Devido ao perfil do espectro, podemos concluir que a proteína seja composta predominante por α -hélices, pois o espectro de proteínas com esse perfil apresentam valores negativos na região de 208 e 222 nm e valores positivos a 193 nm (GREENFIELD, 2006). A análise pelo servidor K2D3 indica que entre a região 190 a 240 nm 56,47% é composta por α hélices e 6,66% por folhas β .



Figura 15– Espectro de dicroísmo circular da proteína TylM2 em 50 mM Tris-hcl, 150 mM NaCl, 10% v/v glicerol, 0,1% v/v Thesit e 5 mM β -Mercaptoetanol, pH 7,8. Perfil do espectro indica que a proteína seja composta predominantemente por α hélices.

4.2.3 Espalhamento de luz dinâmico

Por meio da técnica de espalhamento de luz dinâmico é possível obter estimativas de massa molecular da proteína baseada no raio hidrodinâmico das partículas presente na amostra. Quando um feixe de luz monocromático passa por um polarizador e incide sobre a amostra, toda partícula presente em solução espalha a luz em várias direções e essa luz espalhada passa por um analisador e detector, posicionado em um determinado ângulo Θ e que mede a intensidade das luzes espalhadas. O campo elétrico da luz, ao ser espalhada pelas partículas, terá um efeito sobre as cargas de cada partícula, acelerando-as dentro do solvente (BERNE; PECORA, 2000). Em uma escala de tempo, a flutuação da intensidade de luz espalhada se deve ao fato das moléculas estarem em movimentação constante e aleatória, denominado movimento Browniano. Sua velocidade é definida pelo coeficiente de difusão. Algumas propriedades de partículas que podem influenciar o coeficiente de difusão, e consequentemente a velocidade são: a concentração, tamanho e estrutura superficial das partículas e os tipos de íons que compõem o solvente. Quanto maior for a partícula, mais lento será seu movimento browniano e vice versa. Por padrão, o raio hidrodinâmico medido é de uma esfera hipotética que se difunde na mesma velocidade da partícula de proteína sendo examinada. Fatores que influenciam o raio hidrodinâmico de proteínas são a massa molecular, formato e conformação da proteína.

De acordo com o resultado de DLS, a amostra apresenta duas populações. A primeira população com uma massa molecular de aproximadamente 193 KDa e a segunda população com aproximadamente 2202 KDa. A polidispersividade de uma amostra descreve a distribuição não uniforme das partículas contidas na amostra. Para uma amostra ser considerada monodispersa, a porcentagem de polidispersividade precisa estar abaixo de 20% e somente uma população deve estar presente. A partir dessa análise de DLS para proteína TylM2, foi possível concluir que a amostra estava polidispersa e nenhuma população encontrada é referente ao monômero da proteína TylM2 (50 KDa). Pelo fato de não ter uma população homogênea, o espalhamento múltiplo de luz pode ocorrer no qual a luz espalhada por uma partícula é espalhada por outra antes que chegue ao detector e isso pode comprometer a precisão do cálculo do tamanho da partícula.

Tabela 8- Resultado de espalhamento de luz dinâmico da proteína TylM2

_	População	Raio (nm)	Polidispersividade (%)	Massa (KDa)	Intensidade (%)	Massa (%)
	1	5,640	21,9	193	15,8	78,8
_	2	15,972	24,3	2202	84,2	21,2

4.2.4 Espalhamento de raios-X a baixo ângulo

O espalhamento de raios-x a baixo ângulo (SAXS) é uma técnica bastante útil que gera informações sobre a massa molecular e estado oligomérico de uma proteína em solução. A linearidade da curva Guinier é um bom indicativo da qualidade da amostra sendo que desvios sugerem agregação inespecífica ou interações interparticulares. O programa AutoRg identificou a baixa qualidade dos dados (39%) e análises do comportamento da curva de Guinier para TylM2 (Figura 17a) é indicativo de agregação ou interferências intermoleculares, portanto a amostra não está ideal e com isso não é possível estimar o raio de giro R(g) da proteína com precisão e efetuar a construção do envelope tridimensional. A partir do gráfico de Krakty para TylM2 (Figura 17b) é possível confirmar o enovelamento parcial da proteína, pois após o pico a curva se mantem em valores altos de *q*. O servidor SAXSMoW2 (FISCHER et al., 2010) (disponível em: <u>http://saxs.ifsc.usp.br/</u>), é uma ferramenta utilizada para obter uma estimativa de massa molecular a partir de dados de SAXS. Para TylM2 a massa molecular estimada foi 190 kDa,. Análises de gel-nativo e DLS sugerem que a TylM2 forma um tetrâmero em solução e de certa forma, está em acordo com os valores de massa sugeridos pelo SAXSMoW2.



Figura 16- As curvas de Guinier (a) e Krakty (b) obtidas a partir dos dados de SAXS da glicosiltransferase TylM2.

4.3 Cristalização, Difração e Determinação da Estrutura

4.3.1 Ensaios biofísicos da proteína TylM2

O primeiro cristal para a proteína TylM2 obtido a partir da expressão em *shaker* foi encontrado na condição A5 composta por 3,4 M, 1,6-Hexanediol, 200 mM MgCl₂ e 100 mM Tris-HCl pH 8,5 do kit JBScreen Basic HTS (JenaBioscience), utilizando uma concentração de 10 mg/mL da proteína (Figura 17). As gotas foram preparadas com a proporção de 1 μ L de proteína para 1 μ L da condição de cristalização e em apenas 24 horas após o preparo das placas usando o robô Oryx4 (Douglas Instruments) já foi possível observar a formação de microcristais.



Figura 17- Primeiro monocristal obtido da proteína TylM2 na condição A5- JBScreen Basic HTS (JenaBioscience) composta por 3,4 M 1,6 Hexanediol, 200 mM MgCl₂ e 100 mM Tris-HCl pH 8,5.

Durante o processo de cristalização ocorre um empacotamento dos motivos na condição de cristalização, que é composto por água, moléculas tamponantes, agentes precipitantes, ligantes, cofatores e aditivos. Essa solução normalmente representa 50% do

volume do cristal, formando canais entre os motivos empacotados de forma ordenada (KANTARDJIEFF; RUPP, 2003). No centro desses canais e ao longo do eixo rotacional do plano, as moléculas do solvente encontram-se de forma desordenada enquanto que próximo à camada de coordenação, as moléculas encontram-se mais ordenadas. A densidade eletrônica gerada por moléculas ordenadas do solvente deve ser identificada, pois fará parte do modelo atômico final (WEICHENBERGER et al., 2015).

Cristais, para que difratem bem, devem geralmente apresentar uma forma com arestas bem definidas e possuir um tamanho de 0,5 a 1 μ m. O cristal da proteína TylM2 era pequeno e composto por placas sobrepostas porém, apesar de não apresentar uma forma satisfatória, foi levado para linha MX2 (LNLS-campinas) para ser difratado por raios-X e foi coletado um conjunto de dados, de baixa resolução, com 1800 imagens. Após o processamento inicial dos dados pelo programa XDS, verificou-se que o cristal pertencia ao grupo espacial ortorrômbico P2₁2₁2₁. Os parâmetros utilizados para corte de resolução estão na tabela 9 abaixo.

3 1 2			
Comprimento de onda (Å)	1,46		
Limites de resolução (Å)	3,5-50,05		
Grupo espacial	P2 ₁ 2 ₁ 2 ₁		
Célula unitária (a, b, c e α , β , γ)	136,89 261,40 283,87 90,00 90,00 90,00		
Número de imagens	1800		
Número de reflexões	1707731 (69105)		
Numero de reflexões únicas	128053 (5984)		
$\mathbf{R}_{\mathrm{merge}}$	0.578 (2.532)		
\mathbf{R}_{pim}	0.163 (0.760)		
<i>/<σ(I)></i>	6.9 (1.3)		
Completeza (%)	99.8 (95.9)		
Multiplicidade	13.3 (11.5)		
Moléculas na unidade assimétrica**	40		

Tabela 9- Dados de difração de raios X obtidos para TylM2 no Laboratório Nacional de Luz Sincrotron

*valores entre parênteses são referentes a valores na última camada de resolução.

** Valores baseados no coeficiente de matthews Vm

Para determinar o número de moléculas na unidade assimétrica quando o valor de solvente presente no cristal é de aproximadamente 50%, foi utilizado o coeficiente de Matthews presente no pacote de programas CCP4i e o número estimado de moléculas nessa

porcentagem de solvente foi 40. Este número é extremamente alto e improvável, sugerindo que há patologias como *twinning* ou pseudo-simetria, ou ainda pode indicar uma designação incorreta do grupo espacial (ZWART et al., 2007). Após verificar que o cristal apresentava aproximadamente 40 moléculas na unidade assimétrica, foi utilizado o programa Xtriage presente no pacote Phenix, para determinar se o cristal apresentava patologia e foi identificada uma anormalidade. Isso impediu a continuação do processo de determinação da estrutura terciária da proteína.

A fim de melhorar o aspecto desses cristais, foi realizada uma reprodução manual e otimização da condição utilizando o método de gota pendurada, no qual fatores que promovem o processo de cristalização foram variadas, como a concentração do agente precipitante, pH e temperatura (MCPHERSON; GAVIRA, 2014). A partir do momento em que uma condição de cristalização é encontrada, reduzindo o processo de supersaturação, como diminuindo a concentração da proteína, pode resultar no crescimento mais lento e controlado, que por sua vez poderá gerar cristais morfologicamente melhores (MCPHERSON; CUDNEY, 2014). Mesmo após a tentativa de otimizar os cristais, não foram obtidos novos cristais que fossem significativamente melhores que o original em morfologia e em poder de difração. Foi também preparada uma placa no qual foi adicionado 5 mM de TDP à condição de cristalização na tentativa de estabilizar a proteína e prevenir agregação, e com isso melhorar o cristal porém não houve surgimento de cristais com ligante.

Apesar da proteína TylM2 purificada a partir da expressão em biorreator não apresentar alto grau de pureza, foram montadas placas contendo a mesma condição em que cristais da TylM2 já haviam crescido anteriormente e foram obtidos novos cristais. Esses novos cristais apresentaram uma morfologia muito semelhante à dos primeiros cristais obtidos. Mesmo com a otimização da condição de cristalização, não houve nenhuma melhora expressiva na morfologia dos cristais. Temperatura é outro parâmetro que influencia a supersaturação e a solubilidade da proteína, principalmente se a solução de cristalização apresentar uma força iônica fraca. Dessa forma, condições de cristalização contendo poliálcoois e solventes orgânicos como agente precipitante serão mais sensíveis aos efeitos de temperatura (MCPHERSON; CUDNEY, 2014). O agente precipitante da condição no qual a TylM2 cristaliza é 1,6 Hexanediol, um álcool não volátil, com isso, a fim de melhorar a morfologia do cristal uma nova placa foi preparada e incubada a 4 °C. O processo de

cristalização ocorreu de forma mais lenta, demorando em torno de 3 a 4 semanas e foram obtidos cristais maiores e mais uniformes como pode ser visto na Figura 18.



Figura 18- Imagem de novos monocristais da proteína TylM2 na condição A5- JBScreen Basic HTS (JenaBioscience) composta por 3,4 M 1,6 Hexanediol, 200 mM MgCl₂ e 100 mM Tris-HCl pH 8,5.

Esses cristais foram levados para serem difratados por raio-X na linha BL9-2 (SRRL-USA) e foi coletado um conjunto de dados com 1800 imagens. Apesar de apresentar uma morfologia melhorada, não houve melhora na resolução da difração. Após processamento inicial dos dados pelo programa XDS, foi possível determinar que o cristal fosse monoclínico, desta vez pertencendo ao grupo espacial P2₁. Os parâmetros utilizados para corte de resolução podem ser observados na Tabela 10 abaixo.

Comprimento de onda (Å)	0,97946	
Limites de resolução (Å)	3,65 - 49,49	
Grupo espacial	P12 ₁ 1	
Célula unitária (a, b, c e α , β , γ)	285.2, 135.9, 287.5 90.0, 117.5, 90.0	
Numero de imagens	1800	
Numero de reflexões	618632(31391)	
Numero de reflexões únicas	201487 (10180)	
R _{merge}	0.331 (1.544)	
R_{pim}	0.213 (0.988)	
<i>/<σ(I)></i>	5.4 (1.2)	
Completeza (%)	93.0 (95.3)	
Multiplicidade	3.1 (3.1)	
Moléculas na unidade assimétrica**	40	

Tabela 10- Dados de difração de raios-X obtidos para TylM2 no Stanford Synchrotron Light Source

Curiosamente, o programa Matthews Coeficient do pacote de programas CCP4 indicou novamente um número de aproximadamente 40 moléculas na unidade assimétrica. Com isso, foi realizado outra análise de twinning pelo programa Xtriage (Phenix) e de acordo com o cálculo L, o programa não detectou caso de twinning. Somente em casos de twinning perfeito não é possível determinar a estrutura cristalográfica. Portanto, como não foi detectado esse problema foi iniciado processo de substituição molecular pelo programa phaser para encontrar a fase. Utilizamos a estrutura da glicosiltransferase EryCIII como modelo (código PDB 2YJN), que apresenta uma homologia de 49 % com TylM2.

Foi também utilizado o kit Additive Screen HT (Hampton Research) que consiste em uma biblioteca de 96 tipos diferentes de reagentes que podem estabilizar ou melhorar a conformação da proteína através de interações específicas com a proteína, perturbando interações proteína-proteína ou proteína-solvente, que por sua vez influencia na solubilidade e cristalização da proteína. Após um mês da realização do screeining com aditivos, foram observados o surgimento de micro cristais em duas condições (Figura 19) e após 4 mêses, foi possível verificar que os microcristais haviam, apesar de pouco, crescido mais. Os aditivos G4 e G8 representam 50% v/v polietilenoglicol 400 e 40% v/v polipropilenoglicol P400, respectivamente.



Figura 19- Imagem de cristais obtidos com os aditivos 50% v/v polietilenoglicol 400 e 40% v/v polipropilenoglicol P400, dos poços G4 e G8, respectivamente do Additive Screen HT (Hampton Research).

^{*} valores entre parênteses são referente a valores na última camada de resolução.

^{**} Valores baseado no coeficiente de matthews Vm

4.4 Modelagem molecular e análise de Bioinformática para TylM2 e TylM3

O interesse em predizer estruturas proteicas por métodos *in silico* têm aumentado bastante uma vez que determinar estruturas de proteínas por métodos experimentais nem sempre é bem sucedido. Isso pode ser notado pelo número de sequências primárias conhecidas ser significativamente maior que o número de estruturas tridimensionais determinadas por métodos experimentais. Em julho de 2017 o número de sequências de aminoácidos disponíveis no SWISS-PROT, é de 554.860, enquanto que no PDB, há somente 131.485 estruturas tridimensionais de proteínas disponíveis.

Modelagem molecular utiliza programas computacionais e algoritmos para calcular informações sobre as propriedades de compostos tais como energia mínima, cargas atômicas, entre outras. É uma metodologia que permite construir modelos que possam auxiliar na compreensão das propriedades físico-químicas, características moleculares e investigar a atividade biológica de compostos (BARREIRO et al., 1997). Existem dois métodos para predizer a estrutura tridimensional de proteínas: o primeiro método é a modelagem *ab initio* no qual princípios físicos são aplicados para predição da estrutura da proteína a partir da sequência de resíduos de aminoácidos e sem o uso de um molde (*template*); o segundo método é a modelagem comparativa que é baseada na homologia entre duas ou mais proteínas.

Proteínas derivadas de um ancestral comum, denominadas homólogas, apresentam semelhanças em suas estruturas primárias. Porém, processos genéticos de evolução e mutações resultaram nas divergências entre as sequências de aminoácidos e consequentemente essas divergências fazem com que as proteínas se agrupem em famílias. Desta forma, proteínas que pertencem à mesma família apresentam algumas semelhanças em suas sequências de resíduos de aminoácidos e estruturas tridimensionais (HÖLTJE et al., 2008). Essas regiões conservadas podem ser identificadas através do alinhamento entre sequências. Se há homologia de uma sequência em relação à outra, cuja estrutura já foi caracterizada, de ponto de vista bioquímico e estrutural algumas propriedades e aspectos da estrutura tridimensional podem ser preditos (SANTOS FILHO; ALENCASTRO, 2003).

Devido à dificuldade em obter informações estruturais a partir de dados experimentais, foi realizado a construção de modelos teóricos para as proteínas TylM2 e TylM3. As sequências de cada proteína foram obtidas a partir do banco de dados ENA-EBI (disponível em: http://www.ebi.ac.uk/ena). A sequência de aminoácidos foi utilizada para realizar um alinhamento, auxílio da ferramenta BLAST (disponível com em https://www.ncbi.nlm.nih.gov/) contra todas as proteínas cujas estruturas foram determinadas depositadas no banco de dados de proteínas PDB (disponível e em http://www.rcsb.org/pdb/home/home.do). Os resultados gerados pelo BLAST podem ser vistos na Tabela 11 e Tabela 12 para TylM2 e TylM3, respectivamente. Através do alinhamento, o programa ordena as proteínas que possuem maior grau de similaridade com a proteína alvo. Foram escolhidos apenas proteínas com ID% acima de 30% para servirem de molde na construção dos modelos. Dentre os possíveis moldes, foi escolhida a proteína que apresenta o mesmo mecanismo da TylM2/TylM3, que é a formação de um complexo entre glicosiltransferase e uma proteína auxiliar. A estrutura do código 2YJN é referente ao complexo formado pela glicosiltransferase EryCIII e sua proteína auxiliar EryCII (MONCRIEFFE et al., 2012).

Tuben II Trotemus com estruturus resorvidus que apresentam similaridade a Tym2				
Código PDB	% Identidade	% Cobertura		
2YJN	49	94		
3H41	42	13		
1RRV	42	13		
3WAD	39	92		
4LEI	37	94		
3TSA	30	92		

 Tabela 11- Proteínas com estruturas resolvidas que apresentam similaridade à TylM2

Tabela 12- Proteínas com estruturas resolvidas que apresentam similaridade à TylM3

Código PDB	% Identidade	% Cobertura
5HDI	36	35
2YJN	33	92
4E2P	32	35

Após a construção de 5 modelos para cada proteína, realizado pelo programa MODELLER, a qualidade e confiabilidade dos modelos foram avaliados com os programas PROSA-Web, PROCHECK e VERIFY3D. Além disso, o programa MODELLER gera um valor DOPE que pode ser utilizado como parâmetro para avaliação inicial de cada modelo gerado, no qual o modelo que apresentou o menor valor DOPE foi utilizado para avaliações posteriores pelos programas acima citados. O melhor modelo para a proteína TylM2 apresentou um valor DOPE de -46986.3000781 e o modelo da TylM3, -33860.18359.

O servidor PROSA-Web gera um valor, denominado Z-score, que é indicativo da qualidade geral do modelo. Esse valor é usado para verificar se a estrutura gerada por modelagem comparativa está dentro dos padrões de resultados geralmente encontrados para proteínas com tamanho similar e cujas estruturas foram resolvidas por métodos experimentais como cristalografia de raios-X e ressonância magnética nuclear- RMN. O Z-score para o modelo TylM2 (Figura 20a)foi -9,33 e para o modelo TylM3 (Figura 20b) o z-score foi de – 4,04.



Figura 20- Z scores de todas as proteínas resolvidas por cristalografia de raios X (azul claro) e ressonância magnética nuclear (azul escuro), depositadas no PDB em relação ao tamanho da cadeia polipeptídica. O ponto preto na imagem a-representa os Z-score do modelo Tylm2 e b- o z-score obtido para o modelo Tylm3.

Para a avaliação da qualidade estereoquímica dos modelos gerados foi utilizado o programa PROCHECK que avalia os ângulos torcionais ($\psi \in \varphi$) da cadeia principal e das cadeias laterais (chi); energia de ligações de hidrogênio; impedimentos estéricos; planaridade de ligações peptídicas e desvios em relação a geometria dos Ca. A avaliação média de todos esses parâmetros é representada pelo fator G, onde os cálculos comparativos são baseados em um banco de dados que contem estruturas resolvidas a diferentes níveis de resolução. Valores baixos estão relacionados com uma conformação improvável do modelo e idealmente o valor

deve estar acima de -0,5. Ambos os modelos apresentam valores acima do esperado. O valor da média geral de fatores G para o modelo TylM2 foi -017 e para TylM3, -0,32. Para avaliação de impedimentos estéricos, o programa gera informações do gráfico de Ramachandran (Tabela 13). A estrutura deve apresentar 90% de seus resíduos em regiões favoráveis, desconsiderando glicina, que não possui cadeia lateral, e prolina, cujo C α está ligado à cadeia lateral. O modelo para TylM2 apresentou 89 % dos resíduos em regiões favoráveis, que está bem próximo ao valor ideal. Em contrapartida, o modelo gerado para TylM3 apresentou apenas 80,02% de seus resíduos em regiões favoráveis, indicando que o modelo não está muito confiável e outras estratégias para a construção de um novo modelo pode ser adotada, como a utilização de vários moldes.

Modelo	Região favorável	Adicionalmente favorável	Generosamente favorável	Regiões desfavorávsl
TylM2	89 %	10,2 %	0,5 %	0,3 %
TylM3	80,2 %	14,7 %	2,7 %	2,4 %

 Tabela 13- Valores estatísticos do gráfico de Ramachandran, obtidos pelo programa PROCHECK

A avaliação da qualidade do ambiente químico foi realizada pelo software Verify3D, que mede a compatibilidade de uma dada sequência de aminoácidos e a estrutura tridimensional proveniente dessa sequência; através de um perfil 3D que determina o ambiente químico de cada resíduo, atribuindo um escore referente à uma matriz construída a partir de análises estatísticas do PDB. Idealmente os modelos precisam ter no mínimo 80% de seus aminoácidos com escore acima de 0,2 no perfil 3D-1D. O modelo TylM2 (Figura 21) está dentro dos níveis aceitáveis com 87,39% dos resíduos de aminoácidos com valores acima de 0,2, sendo que nenhum resíduo apresenta um escore abaixo de zero. Já, o modelo gerado para TylM3 (Figura 22) apresenta 86,76% dos resíduos de aminoácidos com valores acima de 0,2 porém a região entre os resíduos S254 e R273 apresenta escore abaixo de zero, indicando uma região problemática do modelo.



Figura 21- Representação do perfil 3D-1D do modelo gerado para TylM2, no qual 87,39% dos resíduos de aminoácidos apresentam um escore acima de 0,2.



Figura 22- Representação do perfil 3D-1D do modelo gerado para TylM3, no qual 86,76% dos resíduos de aminoácidos apresentam um escore acima de 0,2.

Os modelos gerados para TylM2 e TylM3 podem ser visualizados na Figura 23a e Figura 24a, respectivamente. Após uma sobreposição das estruturas secundárias de ambos os modelos com os respectivos moldes (Figura23b e Figura 24b), conseguimos verificar que os modelos gerados se sobrepõem bem, apresentando poucas regiões diferentes.



Figura 23- Modelo da gliosiltransferase TylM2 (a) gerado a partir de modelagem comparativa e a sobreposição entre modelo e molde (b) TylM2 em verde e EryCIII (PDB: 2YJN cadeia A) em azul, respectivamente.



Figura 24- Modelo da gliosiltransferase TylM3 (a) gerado a partir de modelagem comparativa e a sobreposição entre modelo e molde (b) TylM3 em amarelo e EryCII (PDB: 2YJN cadeia B) em roxo, respectivamente.

A literatura têm relatado vários casos de especificidade relaxada de glicosiltransferases tanto por moléculas doadoras quanto aceptoras. De acordo com Jung et al. (2007), a glicosiltransferase TylM2 possui uma especificidade relaxada em relação ao substrato doador porém a especificidade para o substrato aceptor é mais restrita. Para investigar uma possível relação entre o tamanho do sítio ativo com a especificidade, foi utilizado o servidor Pocasa (disponível em: http://altair.sci.hokudai.ac.jp/ g6/service/pocasa/) para comparar o volume das cavidades que compõem o sítio ativo comparando com a EryCIII (Figura 25) . O volume predito da cavidade para TylM2 foi 664 Å³ e a cavidade da homóloga EryCIII foi 1124 Å³. Apesar de não haver estudos investigando a especificidade da EryCIII em relação à diferentes aceptores, a diferença entre o volume da cavidade das duas proteínas é bastante significativo e corrobora a informação obtida experimentalmente por Jung et al. (2007).



Figura 25- Comparação entre o volume da cavidade no qual os substratos se ligam. às glicosiltransferases TylM2 (a) e EryCIII (b), apresentando volumes 664 Å³ e 1124 Å³, respectivamente, e das cargas que compõem a cavidade da glicosiltransferase TylM2 (c), que apresenta predominantemente cargas positivas quando comparado com a cavidade da EryCIII (d), que apresenta uma região com cargas negativas ausente na TylM2.

De acordo com Moncrieffe et al (2012), os aminoácidos envolvidos na ligação do substrato aceptor à glicosiltransferase são: M127, S128, L132, E151, G174, P175, D176, Q183, V222 e V223. Na glicosiltransferase TylM2, esses mesmos resíduos de aminoácidos são representados por: F127, N128, T132, E151, G174, P175, D176, A183, I225 e G226, respectivamente (Figura 26). A maioria desses resíduos são semelhantes, se não iguais, com a exceção de dois: T132 (um aminoácido polar) que em EryCIII é representado pelo aminoácido hidrofóbico L132; e o aminoácido hidrofóbico A183 enquanto que em EryCIII é encontrado o aminoácido polar.



Figura 26- Sobreposição dos resíduos de aminoácidos envolvidos na ligação ao substrato aceptor de TylM2 (em verde) e EryCIII (em azul).

A partir de um mapa de potencial eletrostático é possível verificar a interface de interação entre a TylM2 e sua proteína auxiliar (Figura 27-28). A interface de interação apresenta dois componentes, sendo uma região hidrofóbica e uma região que apresenta cargas complementares que devem estar envolvidas em uma provável interação eletrostática e formação de pontes salinas entre a glicosiltransferase e a proteína auxiliar. Sob análise do complexo EryCIII/EryCII, os resíduos preditos para realizarem pontes de hidrogênio entre TylM3 e TylM2 são respectivamente: L24-E210, R89-A125, R29-N128, Q32-E203, Q32-P200, R35-E198, A41-I86, E76-G88, E76-I86, E76-E85, L115-E81, L214-203, E112-E85 e R402-R196 (Figuras 29 e 30).



Figura 27- Região de interface entre TylM2 (*cartoon* em verde) e TylM3 (superfície eletrostática), com uma visão lateral (a) e frontal com (b) e sem TylM2 (c), onde o quadro preto evidencia a região de interface na TylM3.



Figura 28- Região de interface entre TylM3 (*cartoon* em amarelo) e TylM2 (superfície eletrostática), com uma visão lateral (a) e frontal com (b) e sem TylM3 (c), onde o quadro preto evidencia a região de interface na TylM2.



Figura 29- Interação entre a glicosiltransferase TylM2 (verde) e a proteína auxiliar TylM3 (amarelo) evidenciando os resíduos da glicosiltransferase envolvidos na interação.



Figura 30- Interação entre a glicosiltransferase TylM2 (verde) e a proteína auxiliar TylM3 (amarelo) evidenciando os resíduos da proteína auxiliar envolvidos na interação.

5 CONSIDERAÇÕES FINAIS

Embora não tenha sido possível obter e determinar uma estrutura cristalográfica da glicosiltransferase TylM2 *apo* e/ou em complexo com a proteína auxiliar TylM3, estudos realizados com as homólogas DesVII e DesVIII por Borisova et al. (2008) sugerem que o papel da proteína auxiliar é estabilizar o complexo com a glicosiltransferease. A TylM3 realiza uma mudança conformacional que facilita a reação de glicosilação, já foi demonstrado que na ausência da proteína auxiliar, a glicosiltransferase não apresenta atividade catalítica. A dificuldade em se produzir as proteínas recombinantes pode estar relacionada à ausência do respectivo par das proteínas durante a expressão.

A expressão de proteínas em biorreator se mostrou um método viável, apesar da quantidade de contaminantes produzidos. Cristais da glicosiltransferase foram obtidos na mesma condição de cristalização de ambos os métodos de produção da proteína. Contudo, o grupo espacial em ambos os casos se mostraram diferente. Tanto os cristais obtidos pelo biorreator quanto em *shaker* apresentaram uma predição de 40 moléculas na unidade assimétrica, o que impossibilitou a determinação da estrutura.

O espalhamento de luz dinâmico evidenciou uma polidispersividade da amostra, pois aproximadamente 80% apresentou uma massa molecular referente a um possível tetrâmero. Um resultado semelhante foi obtido a partir do espalhamento de raios-X a baixos ângulos, corroborando com o resultado obtido anteriormente. Apesar do resultado do gel nativo não ser preciso, a presença da banda referente à glicosiltransferase se encontra abaixo de 242 KDa, o que também sugere um estado tetramérico da proteína *in vitro* na ausência da proteína auxiliar.

A modelagem comparativa permitiu gerar um modelo confiável para a proteína TylM2, no qual foi possível obter informações a respeito do sitio ativo. Esse modelo final não representa o estado conformacional no qual a proteína se encontra em complexo com a proteína auxiliar, pois devido ás interações intermoleculares é possível que a glicosiltransferase apresente modificações conformacionais diferente da proteína modelada.
REFERENCIAS*

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–410, 1990.

BAILEY, S. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallographica Section D. v. 50, n. 4, p. 760–763, 1994.

BARREIRO, E. J. et al. Modelagem Molecular: Uma Ferramenta para o Planejamento Racional de Fármacos em Química Medicinal. **Química Nova**, v. 20, n. 3, p. 300–310, 1997.

BERNARD, A; PAYTON, M. Fermentation and growth of Escherichia coli for optimal protein production. **Current protocols in protein science**, Unit5.3, 5.3.1-5.3.18, 2001.

BERNE, B. J.; PECORA, R. Dynamic Light Scattering with Applications to Chemistry, Biology and Physics. USA: Dover Publication, 2000, 384 p.

BORISOVA, S. A et al. Characterization of the glycosyltransferase activity of desVII: analysis of and implications for the biosynthesis of macrolide antibiotics. Journal of the American Chemical Society, v. 126, n. 21, p. 6534–6535, 2004.

BORISOVA, S. A et al. Glycosylation of acyclic and cyclic aglycone substrates by macrolide glycosyltransferase DesVII/DesVIII: analysis and implications. **Chembiochem : a European journal of chemical biology**, v. 9, n. 10, p. 1554–1558, 2008.

BORISOVA, S. A. et al. Substrate specificity of the macrolide-glycosylating enzyme pair DesVII/DesVIII: Opportunities, limitations, and mechanistic hypotheses. Angewandte Chemie - International Edition, v. 45, n. 17, p. 2748–2753, 2006.

BORISOVA, S. A; LIU, H.-W. Characterization of glycosyltransferase DesVII and its auxiliary partner protein DesVIII in the methymycin/picromycin biosynthetic pathway. **Biochemistry**, v. 49, n. 37, p. 8071–8084, 2010.

BOWIE, J. U.; LÜTHY, R.; EISENBERG, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Stucture. **Science**, v. 253, n. 58, p. 164–170, 1990.

BRETON, C. et al. Structures and mechanisms of glycosyltransferases. **Glycobiology**, v. 16, n. 2, p. 29–37, 2006.

BRUNTON, L; LAZO, J; PARKER K. Goodman & Gilman: Farmacological Basis of Therapeutics. 11 ed. USA: McGraw-Hill education, 2010. 1808 p.

^w De acordo com:

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

BURGIE, E. S.; HOLDEN, H. M. Three-Dimensional Structure of DesVI from Streptomyces venezuelae: A Suagar N,N-Dimethyltransferase Required for dTDP-Desosamine Biosynthesis. **Biochemistry**. V.47, p.3982-3988, 2008.

CANE, D. E. Programming of erythromycin biosynthesis by a modular polyketide synthase. **Journal of Biological Chemistry**, v. 285, n. 36, p. 27517–27523, 2010.

CASALI, N. Escherichia coli Host Strains. In: **Methods in Molecular Biology**. v. 235, Totwana, NJ: Humana Press, 2003. 316 p.

CHAN, Y. A et al. Biosynthesis of polyketide synthase extender units. Natural product reports, v. 26, n. 1, p. 90–114, 2009.

CUNDLIFFE, E. Glycosylation of Macrolide Antibiotics in Extracts of *Streptomyces lividans*. **Antimicrobial Agents Chemotherapy**. V. 36, p. 348-352, 1992.

DINOS, G. P.; MICHELINAKI, M.; KALPAXIS, D. L. Insights into the mechanism of azithromycin interaction with an Escherichia coli functional ribosomal complex. **Molecular pharmacology**, v. 59, n. 6, p. 1441–1445, 2001.

ELSHAHAWI, S. I. et al. A comprehensive review of glycosylated bacterial natural products. **Chem. Soc. Rev.**, v. 44, n. 21, p. 7591–7697, 2015.

EVANS, P. R.; MURSHUDOV, G. N. How good are my data and what is the resolution? Acta Crystallographica Section D. v. 69, p. 1204–1214, 2013.

FISCHER, H. et al. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. **Journal of Applied Crystallography**, v. 43, n. 1, p. 101–109, 2010.

GLOSTER, T. M. Advances in understanding glycosyltransferases from a structural perspective. **Current Opinion in Structural Biology**, v. 28, p. 131-141, 2014.

GOLDMAN, R. C., SCAGLIONE, F. The Macrolide-Bacterium Interaction and its Biological Basis. **Current Drug Targets - Infectious Disorders**, v. 4, n.3, p. 241-260, 2004.

GREENFIELD, N. J. Using circular dichroism spectra to estimate protein secondary structure. **Nature Protocols**, v. 1, n. 6, p. 2876–2890, 2006.

HANSEN, J. L. et al. The Structures of Four Macrolide Antibiotics Bound to the Large Ribosomal Subunit. **Molecular Cell**, v. 10, p. 117–128, 2002.

HARMS, È. et al. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. **Nature**, v. 413, n. 6858, p. 814-821, 2001.

HEIA, S. et al. Initiation of polyene macrolide biosynthesis: Interplay between polyketide synthase domains and modules as revealed via domain swapping, mutagenesis, and heterologous complementation. **Applied and Environmental Microbiology**, v. 77, n. 19, p. 6982–6990, 2011.

HÖLTJE, H. D. et al. **Molecular modeling Basic Princilpes and Applications**. 3. ed. Weinheim, Germany: Wiley-VCH, 2008, 228 p.

HONG, J. S. J. et al. The Role of a Second Protein (DesVIII) in Glycosylation for the Biosynthesis of Hybrid Macrolide Antibiotics in Streptomyces venezuelae. Journal of Microbiology and Biotechnology, v. 15, n. 3, p. 640–645, 2005.

ISIORHO, E. A. et al. Structural studies of the spinosyn forosaminyltransferase, SpnP. **Biochemistry**, v. 53, n. 26, p. 4292–4301, 2014.

JUNG, W. S. et al. Bioconversion of 12-, 14-, and 16-membered ring aglycones to glycosylated macrolides in an engineered strain of Streptomyces venezuelae. **Applied microbiology and biotechnology**, v. 76, n. 6, p. 1373–1381, 2007.

KABSCH, W. Integration, scaling, space-group assignment and post-refinement Acta Crystallographica Section D. v. 66, p. 133–144, 2010.

KANTARDJIEFF, K. A.; RUPP, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. **Protein Science**, v. 12, n. 9, p. 1865–1871, 2003.

KHLEBNIKOV, A.; KEASLING, J. D. Effect of lacY expression on homogeneity of induction from the Ptac and Ptrc promoters by natural and synthetic inducers. **Biotechnology Progress**, v. 18, n. 3, p. 672–674, 2002.

KONAREV, P. V et al. PRIMUS : a Windows PC-based system for small- angle scattering data analysis. **Journal of Applied Crystallography**, v. 36, p. 1277–1282, 2003.

LAIRSON, L. L. et al. Glycosyltransferases: Structures, Functions, and Mechanisms. Annual Review of Biochemistry, v. 77, n. 1, p. 521–555, 2008.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1993.

LAWRENCE, A.-M.; BESIR, H. Staining of Proteins in Gels with Coomassie G-250 without Organic Solvent and Acetic Acid. **Journal of Visualized Experiments**, v. 30, p. 2–4, 2009.

LEE, H. Y. et al. Reconstitution and characterization of a new desosaminyl transferase, EryCIII, from the erythromycin biosynthetic pathway. Journal of the American Chemical Society, v. 126, n. 32, p. 9924–9925, 2004.

LIANG, D.-M. et al. Glycosyltransferases: mechanisms and applications in natural product development. **Chemical Society reviews**, v. 44, n. 22, p. 8350–8374, 2015.

LUISI, F. et al. Efeito anti-inflamatório dos macrolideos em doenças pulmonares da infância. **Jornal Brasileiro de Pneumologia**, v. 38, n. 6, p. 786–796, 2012.

LÜTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with threedimensional profiles. **Nature**, v. 356, n. 6364, p. 83–85, 1992.

MARQUES; M. V. **Biologia Moleclar e Genética Bacteriana**. Ribeirão Preto: Sociedade Brasileira de Genética, 2012. 348 p.

MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658–674, 2007.

MCPHERSON, A.; CUDNEY, B. Optimization of crystallization conditions for biological macromolecules. Acta Crystallographica Section F, v. 70, p. 1445–1467, 2014.

MCPHERSON, A.; GAVIRA, J. A. Introduction to protein crystallization. Acta Crystallographica Section F, v. 70, n. 1, p. 2–20, 2014.

MÉNDEZ, C.; SALAS, J. A. Altering the glycosylation pattern of bioactive compounds. **Trends in biotechnology**, v. 19, n. 11, p. 449–456, 2001.

MONCRIEFFE, M. C. et al. Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. **Journal of Molecular Biology**, v. 415, n. 1, p. 92–101, 2012.

MOORE, B. S.; HERTWECK, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. **Natural Product Reports**, v. 19, n. 1, p. 70–99, 2002.

OMURA, S. Macrolide Antibiotics: Chemistry, Biology and Paractice. 2. ed. USA: Academic Press, 2002. 365 p.

PÉREZ, M. et al. PM100117 and PM100118, new antitumor macrolides produced by a marine Streptomyces caniferus GUA-06-05-006A. **The Journal of Antibiotics**, v. 69, p. 388–394, 2015.

PETOUKHOV, M. V et al. ATSAS 2.1 - towards automated and web-supported Small-Angle Scattering data analysis. **Journal of Applied Crystallography**. v. 40, p. 223–228, 2007.

SAMBROOK, J.; RUSSEL, D. W. **Molecular Cloning, a Laboratory Manual**. 3. ed. New York: Cold Spring Harbor Laboratory Press, 2000. 2344 p.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. MODELAGEM DE PROTEÍNAS POR HOMOLOGIA. **Quimica Nova**, v. 26, n. 2, p. 253–259, 2003.

SIPPL, M. J. Recognition of Errors in 3-Dimensional Structures of Proteins. **Proteins-Structure Function and Genetics**, v. 17, n. 4, p. 355–362, 1993.

SIVASHANMUGAM, A. et al. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. **Protein Science**, v. 18, n. 5, p. 936–948, 2009.

STAUNTON, J.; WEISSMAN, K. J. Polyketide biosynthesis: a millennium review. **Natural Product Reports**, v. 18, n. 4, p. 380–416, 2001.

THIBODEAUX, C. J.; MELANÇON, C. E.; LIU, H. Unusual sugar biosynthesis and natural product glycodiversification. **Nature**, v. 446, n. 7139, p. 1008–1016, 2007.

THIBODEAUX, C. J.; MELANÇON, C. E.; LIU, H. Natural-product sugar biosynthesis and enzymatic glycodiversification. **Angewandte Chemie** (International ed. in English), v. 47, n. 51, p. 9814–9859, 2008.

TRUMAN, A. W. et al. Chimeric Glycosyltransferases for the Generation of Hybrid Glycopeptides. **Chemistry and Biology**, v. 16, n. 6, p. 676–685, 2009.

VAGIN, A.; TEPLYAKOV, A. MOLREP: an Automated Program for Molecular Replacement. Journal of Applied Crystallography. v. 30, p. 1022–1025, 1997.

WALSH, C. Molecular mechanisms that confer antibacterial drug resistance. **Nature**, v. 406, n. 6797, p. 775–781, 2000.

WEBB, B.; SALI, A. Comparative protein structure modeling using MODELLER. Current **Protocols in Bioinformatics**. Unit5.6, 5.6.1-5.6.32, 2014.

WEBBER, M. A.; PIDDOCK, L. J. V. The importance of efflux pumps in bacterial antibiotic resistance. **Journal of Antimicrobial Chemotherapy**, v. 51, n. 1, p. 9–11, 2003.

WEICHENBERGER, C. X. et al. The solvent component of macromolecular crystals. Acta Crystallographica Section D, v. 71, p. 1023–1038, 2015.

WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, n.2, p. 407–410, 2007.

WILLIAMS, G. J.; THORSON, J. S. Natural Product Glycosyltransferases: Properties and Applications. Advances in Enzymology and Related Areas of Molecular Biology, v. 76, p. 55–120, 2009.

WITTIG, I.; BRAUN, H.-P.; SCHÄGGER, H. Blue native PAGE. Nature Protocols, v. 1, n. 1, p. 418–428, 2006.

WU, H. et al. In vivo investigation to the macrolide-glycosylating enzyme pair DesVII/DesVIII in Saccharopolyspora erythraea. **Applied Microbiology and Biotechnology**, v. 100, n. 5, p. 2257–2266, 2016.

XUE, Y. et al. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 21, p. 12111–12116, 1998.

YUAN, Y. et al. In vitro reconstitution of EryCIII activity for the preparation of unnatural macrolides. **Journal of the American Chemical Society**, v. 127, n. 41, p. 14128–14129, 2005.

ZHANG, H. et al. The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. **Nature communications**, v. 5, n. 4339, p. 1-12, 2014.

ZWART, P. H. et al. Surprises and pitfalls arising from (pseudo)symmetry. Acta Crystallographica Section D, v. 64, n. 1, p. 99–107, 2007.