

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

GLICOBIOLOGIA ESTRUTURAL E EVOLUÇÃO  
MOLECULAR DA PGLB DE *CAMPYLOBACTER LARI*

Porto Alegre – Brasil

Outubro de 2013

# Glicobiologia Estrutural e Evolução Molecular da PglB de *Campylobacter lari*

*Conrado Pedebos*

Dissertação de Mestrado elaborada no Grupo de Bioinformática Estrutural do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul sob orientação do professor doutor

Hugo Verli

Porto Alegre – Brasil

Outubro de 2013

# Glicobiologia Estrutural e Evolução Molecular da PglB de *Campylobacter lari*

*Conrado Pedebos*

Dissertação submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como parte dos requisitos necessários para a obtenção do grau de Mestre em Biologia Celular e Molecular.

Banca Examinadora:

---

Prof. Hugo Verli  
Centro de Biotecnologia - UFRGS  
**Orientador**

---

Prof. Diego Bonatto  
Centro de Biotecnologia – UFRGS

---

Prof. Hubert Stassen  
Instituto de Química – UFRGS

---

Prof. Nelson J. Rosa Fagundes  
Departamento de Genética - UFRGS

---

Dr. Laércio Pol-Fachin (suplente)  
Pesquisador colaborador do Grupo de Bioinformática Estrutural - UFRGS

Esta dissertação foi desenvolvida sob a orientação do Professor Doutor Hugo Verli, com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), da Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS) e da Pró-Reitoria de Pós-Graduação (UFRGS), como requisito para obtenção do grau de Mestre em Biologia Celular e Molecular, junto ao Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul.

# FICHA CATALOGRÁFICA

## CIP - Catalogação na Publicação

Pedebos, Conrado

Glicobiologia Estrutural e Evolução Molecular da PglB de *Campylobacter lari* / Conrado Pedebos. -- 2013.

89 f.

Orientador: Hugo Verli.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia do Estado do Rio Grande do Sul, Programa de Pós-Graduação em Biologia Celular e Molecular, Porto Alegre, BR-RS, 2013.

1. Biologia Estrutural. 2. Biologia Computacional. 3. Oligossacariltransferase. 4. Glicosilação. 5. Filogenia. I. Verli, Hugo, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

## AGRADECIMENTOS

Primeiramente, agradeço ao Prof. Hugo Verli por todo o tempo dedicado a minha orientação desde a iniciação científica. Por sempre trazer novas ideias, por resolver os problemas que enfrentei durante esse período e por motivar, tanto em termos de produção como de inspiração, o meu crescimento como pesquisador. Por seus diversos conselhos, ensinamentos e disposição em prover auxílio sempre que necessário, bem como pela atenção dada a cada detalhe relacionado ao trabalho desenvolvido pelos seus alunos.

Agradeço aos colegas e amigos do Grupo de Bioinformática Estrutural pela convivência e companheirismo. Por dividirmos tanto momentos de tensão como de descontração. Particularmente, agradeço ao Dr. Laércio Pol-Fachin, que me coorientou por alguns anos durante a iniciação científica, sendo meu guia pelas diferentes metodologias desenvolvidas no laboratório e participando ativamente da minha formação. Ainda, por partilharmos tanto conhecimento como publicações.

À UFRGS e, em especial, ao Centro de Biotecnologia por serem centros de excelência, possibilitando a realização de pesquisas científicas do mais alto nível. Também agradeço ao Programa de Pós-Graduação em Biologia Celular e Molecular por proporcionar uma formação de qualidade e pela confiança no trabalho realizado. Aproveito para agradecer também a Secretaria do PPGBCM que sempre resolve as questões com presteza e agilidade, facilitando a vida dos alunos e professores.

Aos membros da banca, por aceitarem o convite.

Agradeço aos meus pais por sempre priorizarem meu desenvolvimento humano e intelectual. Pela constante dedicação, carinho e preocupação, assim como por todo o suporte nas minhas escolhas. Por não pouparem esforços, sempre visando o meu bem-estar.

Agradeço à Soraia Da Cás por me apoiar constantemente e por sempre me incentivar em todos os aspectos da vida. Pela amizade, amor, confiança e companheirismo. Por ser alguém que me inspira a buscar o meu melhor sempre.

*“Science is facts; just as houses are made of stone, so is science made of facts; but a pile of stones is not a house, and a collection of facts is not necessarily science.”*

**Jules Henri Poincaré (1854-1912)**

## SUMÁRIO

LISTA DE ABREVIATURAS.....	ix
RESUMO .....	x
ABSTRACT.....	xi
ÍNDICE DE FIGURAS.....	xii
1. INTRODUÇÃO.....	1
1.1 GLICOSILAÇÃO .....	1
1.2 N-GLICOSILAÇÃO.....	3
1.3 N-GLICOSILAÇÃO EM BACTÉRIAS .....	6
1.4 ASPECTOS ESTRUTURAIS E FILOGENÉTICOS DAS OSTs.....	10
1.5 APLICAÇÕES BIOTECNOLÓGICAS E A GLICOENGENHARIA DE PROTEÍNAS .....	15
1.6 DINÂMICA MOLECULAR.....	17
2. OBJETIVOS.....	20
3. MÉTODOS.....	21
3.1 PROGRAMAS UTILIZADOS.....	21
3.2 PREPARAÇÃO DO SISTEMA.....	21
3.3 INSERÇÃO EM MEMBRANA .....	22
3.4 DINÂMICA MOLECULAR.....	23
3.5 INFERÊNCIA FILOGENÉTICA.....	24
4. RESULTADOS.....	25
4.1 STRUCTURAL AND DYNAMICS ASSESSMENT OF THE BACTERIAL OLIGOSACCHARYLTRANSFERASE PGLB .....	26
4.2 FILOGENIA .....	49
5. DISCUSSÃO GERAL .....	53
6. CONCLUSÕES .....	59
7. PERSPECTIVAS.....	60
8. REFERÊNCIAS BIBLIOGRÁFICAS .....	61
9. APÊNDICES .....	70
10. CURRICULUM VITÆ .....	75

## LISTA DE ABREVIATURAS

DGC	Distúrbios Congênitos da Glicosilação
OST	Oligossacariltransferase
Und-P	Undecaprenila fosfatada
Und-PP	Undecaprenila difosfatada (pirofosfatada)
GalNAc	N-acetilgalactosamina
GlcNAc	N-acetilglicosamina
Glc	Glicose
CC	<i>Central Core</i> (Núcleo Central)
IS	<i>Insertion Sequence</i> (Sequência de Inserção)
P1	Periferia 1
P2	Periferia 2
EL	<i>External loop</i> (Alça externa)
TM	Transmembrana (domínio da PglB)
PP	Periplasmático (domínio da PglB)
GTases	Glicosiltransferases
SPC	Single point charge
BLAST	Basic Local Alignment Search Tool
POPE	1-palmitoil-2-oleil- <i>sn</i> -glicero-3-fosfoetanolamina
LPS	Lipopolissacarídeo
DM	Dinâmica Molecular
PME	Particle Mesh Ewald
Bac	Bacilosamina
PDB	Protein Data Bank
MV	Máxima Verossimilhança
EM	Evolução Mínima
IB	Inferência Bayesiana

## RESUMO

A N-glicosilação é um processo pós-traducional que se caracteriza pela adição de um oligossacarídeo a cadeia lateral de um resíduo de asparagina, contida na assinatura D/E-X-N-X-S/T em bactérias. Na última década, a partir do organismo modelo *Campylobacter jejuni* identificaram-se diversas enzimas participantes da rota da N-glicosilação. Dentre estas, a proteína responsável pela transferência em bloco da cadeia oligossacarídica para a proteína aceptora é a oligossacariltransferase PglB, cuja estrutura foi obtida por cristalografia em 2011. A partir destes avanços, o presente trabalho visa contribuir no entendimento do processo de N-glicosilação ao acrescentar à PglB cristalografada a presença de explícita de solvente, flexibilidade e substratos. Ainda, busca caracterizar filogeneticamente a proteína compreendendo os três domínios da vida, descrevendo a sua história evolutiva. Os resultados obtidos indicam que a PglB pode possuir diferentes conformações para o sítio catalítico, algo não observado na estrutura cristalográfica, alternando entre esses estados durante o tempo. Além disso, a proteína demonstrou realizar movimentos que sugerem a existência de uma espécie de alosteria no sítio catalítico, onde a proteína aceptora glicosilada tem sua saída facilitada a partir de alterações conformacionais comunicadas entre as cavidades da enzima. O motivo conservado WWDYGY manteve-se fortemente interagindo com o substrato, enquanto o motivo MIV pouco interagiu com a treonina da posição +2. Dessa maneira, os dados desse trabalho auxiliam no entendimento da glicobiologia estrutural da PglB e trazem novas informações relativas a catálise. Adicionalmente, dados preliminares sobre a ancestralidade dessa enzima formam a primeira amostragem de uma árvore compondo os três domínios, assim como uma segunda árvore descreve a transição do motivo MIV no domínio Bacteria.

## ABSTRACT

N-glycosylation is a post-translational feature which is characterized by the addition of an oligosaccharide to the side-chain of an asparagine residue presented in the motif D/E-X-N-X-S/T in Bacteria. In the last decade, studies in the *C. jejuni* species identified many enzymes participating in the N-glycosylation pathway. Among these, the protein responsible for the *en bloc* transference of the oligosaccharide chain to the protein acceptor is the oligosaccharyltransferase PglB, whose structure was obtained in 2011 by x-ray crystallography. Departing from these findings, the present work aims to contribute for the comprehension of the N-glycosylation process by adding explicit solvent, flexibility and substrates to the PglB crystallographic structure. Still, we attempt to infer the phylogeny of the protein comprising the three domains of life, describing the evolutionary history of the enzyme. The obtained results indicate that PglB may possess different conformations for the catalytic site, a characteristic not observed in the crystal structure, alternating between states. Besides, the protein demonstrated movements that suggest the existence of an allosteric mechanism in the catalytic site, where the exit of the glycosylated protein is facilitated by conformational transitions communicated between cavities. The conserved motif WWDYGY performed strong interactions with the proteic substrate, while the MIV motif did not. Thus, these data may contribute to understand the structural glycobiology of PglB and bring new insights regarding the catalysis mechanism. Additionally, preliminary data involving the ancestry of this enzyme demonstrated the first phylogenetic tree comprising the three domains, as well as a second tree which describes the transition of the MIV motif in the Bacteria domain.

## ÍNDICE DE FIGURAS

Figura 1. Os tipos de glicosilação identificados até o presente momento e os resíduos de aminoácidos envolvidos. ....	2
Figura 2. Vias da N-glicosilação presentes em três espécies de diferentes domínios da vida: A) <i>C. jejuni</i> (Bacteria); B) <i>M. voltae</i> (Archaea) and C) <i>S. cerevisiae</i> (Eukarya) (Adaptado de Larkin & Imperiali, 2011).4	4
Figura 3. Distribuição dos tipos de glicanas detectadas até o momento para as espécies do domínio Bacteria, sendo que no Grupo I estão presentes as termotolerantes mais o hexassacarídeo de <i>C. lari</i> e, no Grupo II, as não-termotolerantes (Adaptado de Nothaft et al., 2012). ....	8
Figura 4. Organização das unidades estruturais já encontradas para o domínio periplasmático das OSTs espalhadas pelos três domínios. ....	11
Figura 5. Estrutura da enzima PglB cristalizada juntamente com um peptídeo acceptor contendo a assinatura para N-glicosilação. ....	13
Figura 6. Estrutura do sítio catalítico da PglB de <i>C. lari</i> (Lizak et al., 2011). Uma rede de ligações de hidrogênios ocorreria juntamente à manutenção do íon divalente $Mg^{2+}$ no centro, que por sua vez coordenaria tanto as cadeias laterais dos resíduos catalíticos como o lipídio carreador. ....	14
Figura 7. Representação esquemática da engenharia envolvida na expressão de N-glicosilação em <i>E. coli</i> . (adaptado de Baker et al., 2013). ....	17
Figura 8. Potenciais envolvidos na descrição dos termos do campo de força que será empregado em simulações por dinâmica molecular. ....	18
Figura 9. Árvore filogenética envolvendo os três domínios da vida e obtida pelo método de EM. Destacadas estão as bactérias (violeta), archaeas metanógenas (ciano), euryarcheota (verde), crenarcheota (amarelo) e eucariotos (vermelho). ....	50
Figura 10. Árvore filogenética envolvendo os três domínios da vida e obtida pelo método de ML. Destacadas estão as bactérias (violeta), archaeas metanógenas (ciano), euryarcheota (verde), crenarcheota (amarelo) e eucariotos (vermelho). ....	51
Figura 11. Árvore filogenética obtida pelo método de IB e compreendendo espécies do domínio Bacteria e três sequências de grupo externo. Os grupos estão coloridos de acordo com a variante do motivo MIV apresentado. Em dois casos não foi possível identificar qual variação os organismos apresentavam. ....	52
Figura 12. Descrição das interações de motivos conservados com aminoácidos importantes para o processo de N-glicosilação, identificadas a partir das simulações de DM. ....	56

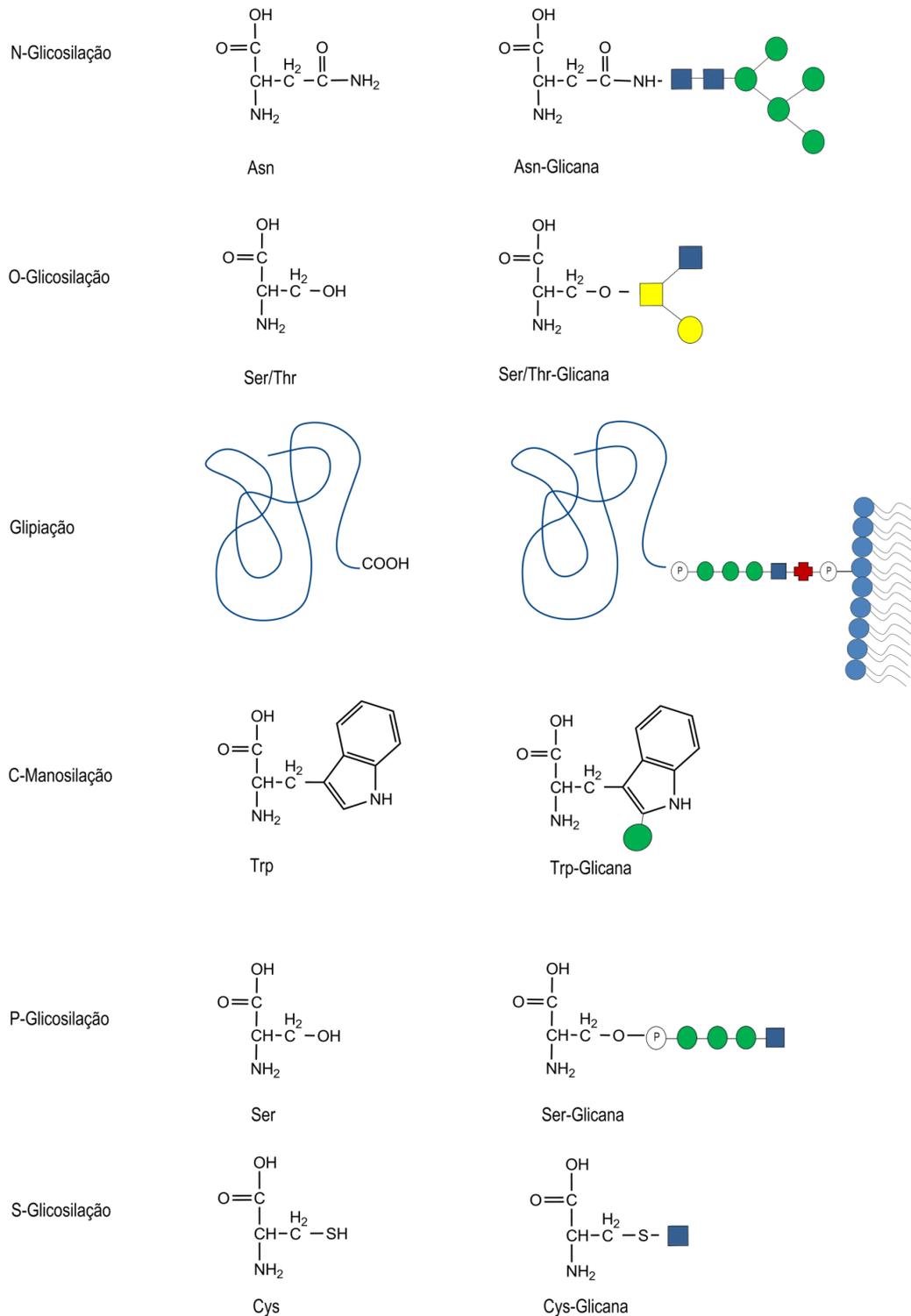
# 1. INTRODUÇÃO

## 1.1 GLICOSILAÇÃO

A glicosilação é uma das modificações co- ou pós-traducionais mais prevalentes já descritas (Varki, 1993; Spiro, 2002) sendo que cerca de dois terços das proteínas presentes em eucariotos possuem potencialmente essa característica (Apweiler *et al.*, 1999). As primeiras evidências da presença desse processo em organismos data do início do século XX, compreendendo estudos realizados por Albert Neuberger que envolviam a ovalbumina cristalizada e a detecção de resíduos de manose e glicosamina (Neuberger, 1938). A partir desse estudo, foi obtida a prova irrefutável da existência de carboidratos ligados à estrutura de proteínas. Desde então, uma vasta gama de funções foram relacionadas à presença dessas moléculas em proteínas como na termodinâmica de enovelamento de proteínas (Shental-Bechor & Levy, 2008), na sinalização celular, na localização intracelular, na resposta imunitária e na estabilidade contra proteólise, entre outros (Varki, 1993). Os processos de glicosilação estão diretamente envolvidos em diversas patologias observadas na prática clínica (Schachter & Freeze, 2009), como os chamados Distúrbios Congênitos da Glicosilação (*Congenital Disorders of Glycosylation - CDG*) que já somam mais de 60 tipos diferentes descritos na literatura (Jaeken, 2011) e atingem as vias responsáveis pela biossíntese de glicoconjugados.

São encontrados diversos tipos de glicosilação na natureza (Figura 1): a O-glicosilação envolve a adição de carboidratos a um átomo de oxigênio de um resíduo de aminoácido, geralmente uma serina ou uma treonina; na P-glicosilação, que envolve uma serina fosfatada, o átomo de fósforo será o acceptor da glicana; a S-glicosilação envolve a adição de monossacarídeos a cadeia lateral de uma cisteína, a partir do átomo de enxofre ali presente; a C-manosilação está relacionada à presença de açúcares em um átomo de carbono do anel indólico de um resíduo de triptofano; a Glipiação é a glicosilação de um lipídio chamado âncora de glicosilfosfatidilinositol (GPI), podendo ser adicionado à porção C-terminal de uma proteína para que essa seja ancorada a uma membrana celular e, por fim, a N-glicosilação é

caracterizada por uma ligação do tipo  $\beta$ -glicosilamida de uma cadeia oligossacarídica a um átomo de nitrogênio presente na cadeia lateral de uma asparagina aceptora (Spiro, 2002; Stepper *et al.*, 2011).



**Figura 1.** Os tipos de glicosilação identificados até o presente momento e os resíduos de aminoácidos envolvidos.

## 1.2 N-GLICOSILAÇÃO

A N-glicosilação é uma das formas de glicosilação de maior ocorrência e está presente nos três domínios da vida (Nothaft & Szymanski, 2010; Calo *et al.*, 2010). Primeiramente encontrada em Eucariotos, a partir de estudos realizados novamente com a ovalbumina por Neuberger e colaboradores (Johansen *et al.*, 1961), apenas recentemente esses sistemas foram descobertos em bactérias (Szymanski *et al.*, 1999, Young *et al.*, 2002), o que aponta para o maior conhecimento da via eucariótica, principalmente através do modelo *Saccharomyces cerevisiae*, além de ser conservada entre os organismos pertencentes a esse domínio (Lehle *et al.*, 2006). Todavia, na última década, a massificação do sequenciamento genômico de procariotos, aliado ao aperfeiçoamento das técnicas de espectrometria de massas, propiciou a identificação da presença maciça dos sistemas de N-glicosilação em diversas espécies de bactérias e *archaeas* (Larkin & Imperiali, 2011).

Uma das poucas vias da N-glicosilação em bactérias bem compreendida e elucidada até o momento ocorre em *Campylobacter jejuni* (Nothaft & Szymanski, 2013) e, em menor grau, as vias existentes em *archaeas*, sobretudo devido à alta complexidade e variabilidade das cadeias de oligossacarídeos. Ainda assim, os passos iniciais pertencentes a esse processo, relativos à montagem da glicana e adição dessa a proteína aceptora, são bastante semelhantes em todos os níveis (Figura 2).

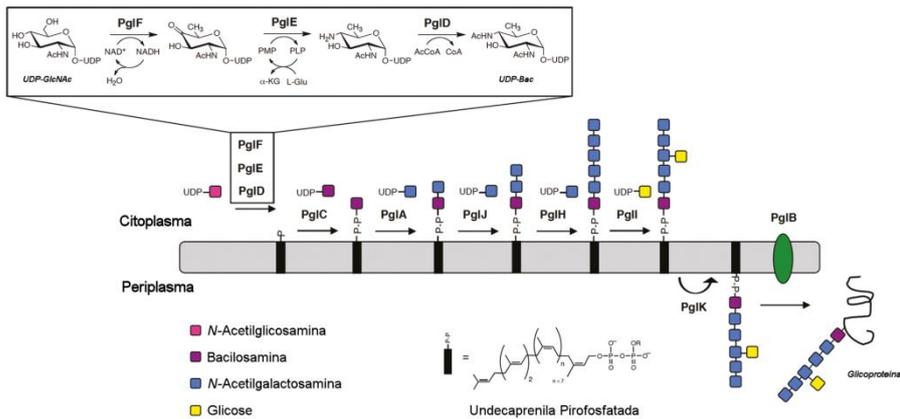
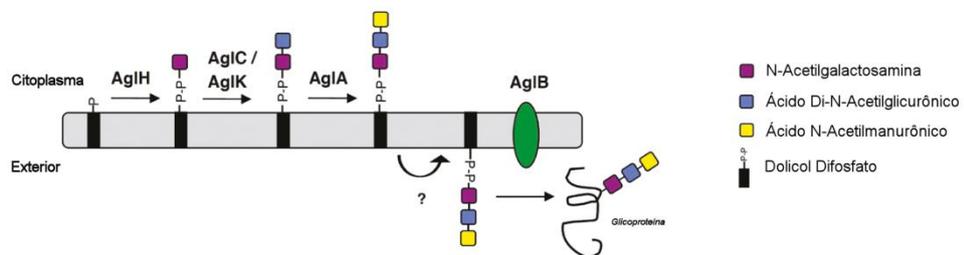
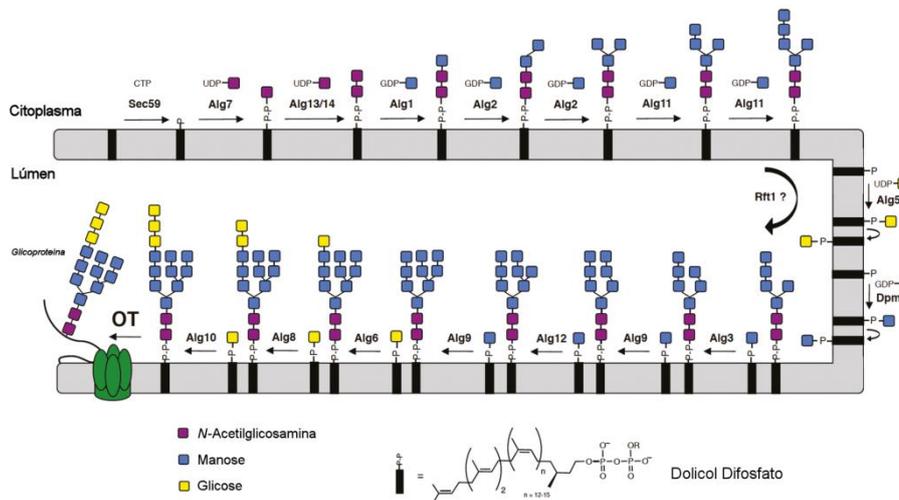
A) Bacteria (*C. jejuni*)B) Archaea (*M. voltae*)C) Eukarya (*S. cerevisiae*)

Figura 2. Vias da N-glicosilação presentes em três espécies de diferentes domínios da vida: A) *C. jejuni* (Bacteria); B) *M. voltae* (Archaea) and C) *S. cerevisiae* (Eukarya) (Adaptado de Larkin & Imperiali, 2011).

Em geral, eles envolvem a presença de diversas enzimas que irão participar na montagem do oligossacarídeo, o qual é transportado via um carreador lipídico. Esse carreador pode ser uma molécula de undecaprenila difosfatada (Und-PP),

em bactérias, ou um dolicol, tanto mono como difosfatado, em *archaeas* e eucariotos. Em relação às enzimas da via (Larkin & Imperiali, 2011), as responsáveis pela montagem do substrato sacarídico em eucariotos são as glicosiltransferases da família *asparagine-linked glycosylation* (Alg) as responsáveis pela montagem do substrato sacarídico, adicionando os monossacarídeos um a um ao carreador. Similarmente, em *archaeas*, as enzimas do tipo glicosiltransferases da família *archaeal glycosylation* (Agl) estão envolvidas no processo, enquanto em bactérias as responsáveis são as da família *protein glycosylation* (Pgl). Essas enzimas utilizam monossacarídeos fosfatados, advindos de açúcares das vias da pentose-fosfato, como substrato para a construção da cadeia oligossacarídica. Toda essa linha de montagem irá ocorrer com o carreador lipídico voltado para o citoplasma tanto nas células procarióticas como nas eucarióticas. Ainda, é importante ressaltar que diferentes enzimas de uma mesma família estarão envolvidas na montagem dependendo do tipo de carboidrato que a determinada espécie produz.

Para que possa ocorrer a transferência da glicana é necessário que esse carreador esteja presente no compartimento celular apropriado: em eucariotos, a adição da glicana à proteína se dá no lúmen do retículo endoplasmático; em *archaeas*, acontece entre a parede celular e a membrana plasmática; já em bactérias, o processo termina no espaço periplasmático, entre a membrana interna e a externa. Para que o oligossacarídeo ligado ao carreador lipídico esteja ao alcance das enzimas denominadas oligossacariltransferases (OST), é necessária a ação de enzimas chamadas flipases, que realizam a translocação desse carreador glicosilado a partir da realização do movimento de flip-flop. Em eucariotos, acreditava-se até pouco tempo que a proteína Rft1 era a responsável por essa função (Helenius *et al.*, 2002), embora essa hipótese tenha sido refutada. Após novas investigações, ela foi implicada apenas como participante no processo (Jelk *et al.*, 2013). Em *archaeas*, na via da espécie *Haloferax volcani*, foi detectado que a enzima AglR pode desempenhar esse papel (Kaminski *et al.*, 2012), enquanto em bactérias a enzima PglK é designada como a flipase envolvida (Kelly *et al.*, 2006).

Após a translocação do carreador glicosilado, a adição de monossacarídeos à cadeia pode continuar (como ocorre em eucariotos e em algumas *archaeas*) ou cessar (como ocorre em bactérias). Com a glicana

pronta, o passo final é a transferência dessa para a proteína a ser glicosilada. Para tanto, será necessária a ação das OSTs, enzimas catalisadoras desse evento. Elas podem variar bastante de tamanho de espécie para espécie e também em termos de características estruturais. As OSTs eucarióticas constituem-se em um complexo composto de 5 a 8 subunidades, incluindo a Stt3, que possui a função catalítica (Dempski Robert & Imperiali, 2002; Kelleher & Gilmore, 2006). Já as *archaeas* e bactérias possuem apenas uma subunidade, a AgIB (Yurist-Doutsch *et al.*, 2008) e a PglB (Wacker *et al.*, 2002), respectivamente, semelhante a Stt3 de eucariotos. Finalizado o processo de transferência, a proteína pode manter-se com a glicana na forma que foi transferida ou essa árvore glicídica pode ainda sofrer alterações adicionais, promovidas no retículo endoplasmático e no Complexo de Golgi (eucariotos) ou por enzimas adicionais (*archaea*), descritas como processadoras da glicana (Herscovics, 1999; Roth, 2002). Em bactérias, esse tipo de modificação ainda não foi detectado; entretanto, em eucariotos, essa é a principal fonte de diversidade estrutural para essas cadeias de açúcares.

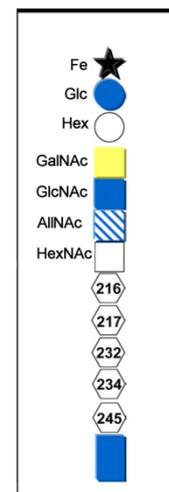
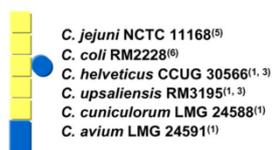
### 1.3 N-GLICOSILAÇÃO EM BACTÉRIAS

As espécies da classe Epsilonproteobacteria são as que mais apresentam genes que codificam o sistema de N-glicosilação (Nothaft & Szymanski, 2010, Nothaft & Szymanski, 2013). Dentro dessa classe, destacam-se os gêneros *Campylobacter*, *Helicobacter* e *Wolinella*. Em menor proporção, alguns membros da classe Deltaproteobacteria, como o gênero *Desulfovibrio*, também possuem o já citado sistema (Ielmini & Feldman, 2011). Outrossim, um sistema diferente foi descrito na classe Gamaproteobacteria – o da GTase HMW1C, cuja atividade é desempenhada no citoplasma ao glicosilar a adesina HMW1 (Grass *et al.*, 2010). Essa OST parece ter evoluído de forma independente, tanto estruturalmente, por possuir pouca similaridade com a PglB, como em relação a localização, por estarem em compartimentos celulares diferentes (Nothaft & Szymanski, 2013).

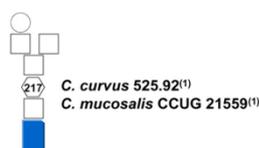
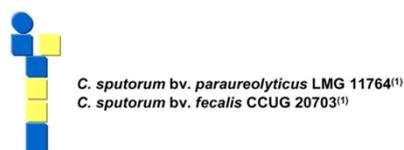
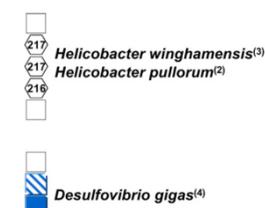
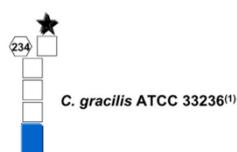
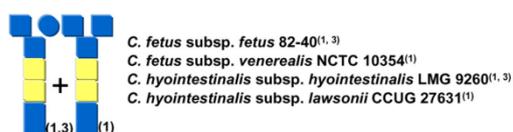
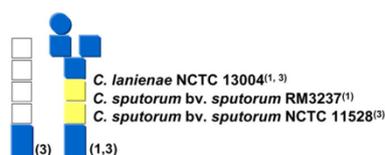
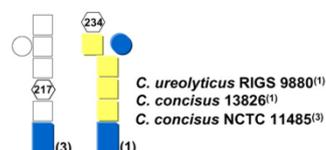
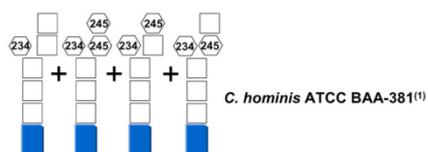
A variabilidade de estruturas sacarídicas N-ligadas é um dos temas que vem sendo constantemente estudados. Nothaft e colaboradores (Nothaft *et al.*, 2012) classificaram em 2 grupos: o Grupo I, ou Grupo Cj, é composto por espécies de linhagens termotolerantes que produzem o heptassacarídeo

presente em *C. jejuni* (Figura 3A). As exceções do grupo são as espécies que possuem a glicana de *C. lari*, um hexassacarídeo com estrutura linear igual a de *C. jejuni*, apenas faltando a glicose da ramificação; o outro grupo é o denominado Grupo II, e compreende as espécies não-termotolerantes da ordem *Campylobacterales* (Figura 3B). Esse grupo apresenta glicanas que podem conter de 6 a 8 monossacarídeos, inclusive com variações estruturais nesses açúcares dentro de uma mesma espécie. Grande parte desses oligossacarídeos ainda não foi elucidada, e possui modificações adicionais como a presença de fosfoetanolamina em algumas das cadeias, o que contribui para a falta de informações sobre as diferentes vias encontradas no domínio Bacteria.

## Grupo I (Termotolerantes)



## Grupo II (Não-termotolerantes)



**Figura 3.** Distribuição dos tipos de glicanas detectadas até o momento para as espécies do domínio Bactéria, sendo que no Grupo I estão presentes as termotolerantes mais o hexassacarídeo de *C. lari* e, no Grupo II, as não-termotolerantes (Adaptado de Nothaft *et al.*, 2012).

A característica mais conservada entre essas glicanas é a presença da N,N'-diacetilbacilosamina ou Bacilosamina (Bac), um monossacarídeo relativamente comum em bactérias e que está diretamente envolvido na ligação com a Asn. Pode se dizer que a via da N-glicosilação em *C. jejuni* inicia a partir da conversão de uma N-acetilglicosamina (GlcNAc) em uma Bac. As enzimas envolvidas nesse processo são a PglF, PglE e PglD, e todas irão agir na posição 4 do monossacarídeo GlcNAc. Primeiramente, a PglF, uma desidratase, retira o átomo de hidrogênio (H) do grupo hidroxila (OH), a partir

de uma molécula de  $\text{NAD}^+$ , que se tornará NADH. Subsequentemente, a aminotransferase PglE irá adicionar o grupo amina, removendo o grupo cetona ali presente (Schoenhofen *et al.*, 2006). Finalizando o processo, a PglD promove a catálise da acetilação, a partir de acetil-CoA (Olivier *et al.*, 2006), finalizando a formação da Bac. Essa, por sua vez, é encontrada na forma ativada UDP-Bac, possuindo um grupo pirofosfato conectado na posição 1, seguido por uma ribose e uma base nitrogenada (uracila). A PglC é uma fosfoglicosiltransferase e age na transferência da Bac fosforilada na posição 1 (Bac-1-P) para o seu carreador lipídico (Und-P) (Glover *et al.*, 2006). Consecutivamente, as PglA e J (Weerapana *et al.*, 2005; Glover *et al.*, 2005) adicionam a primeira e a segunda GalNAc com ligações  $\alpha$ -(1 $\rightarrow$ 3) e  $\alpha$ -(1 $\rightarrow$ 4). A PglH é a responsável pela adição dos próximos três monossacarídeos GalNAc ligados  $\alpha$ -(1 $\rightarrow$ 4) (Troutman & Imperiali, 2009), sendo considerada uma polimerase. O passo final da montagem é a geração da ramificação, a partir da ação da PglI que irá adicionar um resíduo de glicose (Glc) com uma ligação  $\beta$ -(1 $\rightarrow$ 3) (Glover *et al.*, 2005).

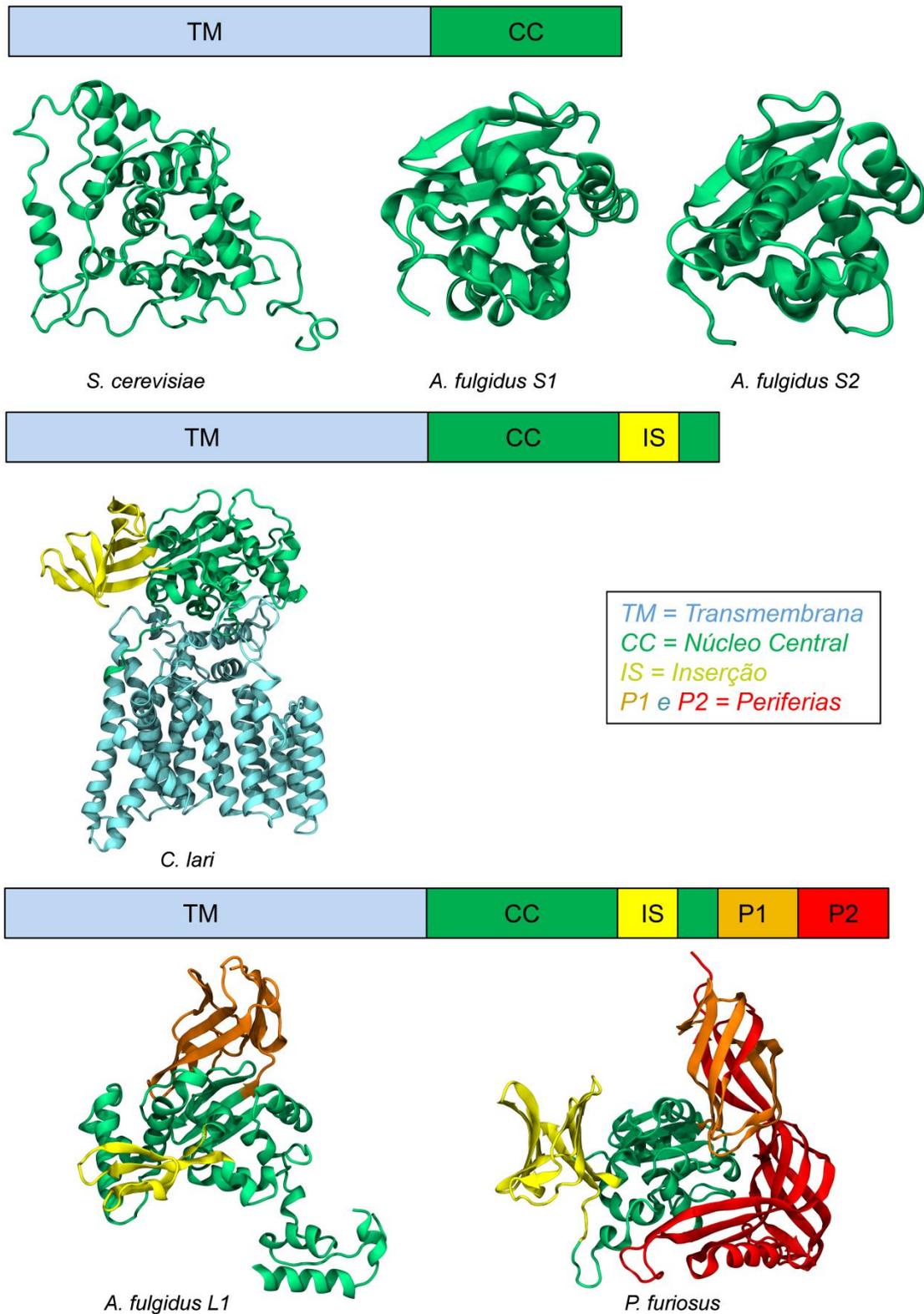
Todo o processo realizado na linha de montagem do heptassacarídeo culminará com a ação da PglB, a OST de bactérias. Para tal fim, a proteína aceptora da glicosilação necessita possuir uma assinatura em sua sequência. A assinatura padrão para N-glicosilação nos domínios da vida é composta pelo motivo N-X-S/T, onde N é a asparagina que receberá o carboidrato, X corresponde a qualquer aminoácido exceto prolina e S/T corresponde a um resíduo de serina ou treonina. Todavia, a PglB demanda uma assinatura estendida para obter maior eficiência (Kowarik *et al.*, 2006), apesar desse requerimento não ser obrigatório para a ocorrência da modificação pós-traducional (Schwarz *et al.*, 2011). Em bactérias, o motivo é composto por D/E-X-N-X-S/T, onde os dois novos resíduos de aminoácido incluídos correspondem a um aspartato ou um glutamato e qualquer aminoácido exceto prolina. Além da presença de uma assinatura estendida, como esse processo ocorre pós-traducionalmente e com a proteína enovelada (Kowarik *et al.*, 2006), diferentemente do processo co-traducional em eucariotos, a região desse motivo precisa estar em uma alça flexível e exposta para que, assim, possa interagir com o sítio catalítico da PglB.

#### 1.4 ASPECTOS ESTRUTURAIS E FILOGENÉTICOS DAS OSTs

Os estudos de cunho estrutural envolvendo as OSTs são relativamente recentes e, até 2011, possibilitavam apenas a observação de mecanismos relacionados à porção localizada no periplasma celular. Isto deve-se a localização das OSTs em membrana, o que torna a caracterização cristalográfica de estruturas completas bastante complexa. Diversas estruturas da porção externa já foram depositadas no banco de dados do Protein Data Bank (PDB), incluindo aquelas das espécies *C. jejuni*, *Archaeoglobus fulgidus* (3 isoformas), *Pyrococcus furiosus*, *Pyrococcus horikoshii* e *S. cerevisiae*. A única estrutura completa de OST é a presente em *C. lari*, a PglB, composta por 712 resíduos com 13 hélices transmembrana.

Partindo dessas estruturas disponíveis, diferentes organizações para o domínio periplasmático foram identificadas, sendo essa região dividida e classificada em unidades estruturais (Igura *et al.*, 2008, Maita *et al.*, 2010) (Figura 4). A estrutura presente em todas as espécies é o núcleo central (*central core* – CC), unidade onde estão localizados os principais motivos conservados dessa enzima, como a sequência WWDXGX, que funciona no reconhecimento da posição +2 da assinatura (S/T) (Lizak *et al.*, 2011). Em geral, é disposta por várias  $\alpha$ -hélices e algumas fitas- $\beta$ , em um enovelamento do tipo  $\alpha/\beta$ . Outra unidade identificada é a sequência de inserção (*insertion sequence* – IS), denominada dessa forma pois parece ter sido de fato inserida dentro da sequência do CC. Ela ocorre em Bacteria e em Archaea, sendo que nas estruturas elucidadas até agora essa unidade apresentou-se com estrutura secundária formada por diversas fitas- $\beta$ , sem um padrão definido de enovelamento. As primeiras análises sugeriram envolvimento desta região no reconhecimento do substrato sacarídico (Maita *et al.*, 2010), todavia a estrutura resolvida da PglB nega essa possibilidade, como será mostrado mais adiante (Lizak *et al.*, 2011). Por fim, ocorrem as unidades nomeadas de periferias 1 e 2 (*periphery* 1 e 2 - P1 e P2). Essas estruturas peculiares formam grandes folhas- $\beta$ , envolvendo parte do CC, na região mais externa da proteína. Estudos (Matsumoto *et al.*, 2012) demonstraram que essas unidades não são essenciais para a catálise da N-glicosilação, contudo, acredita-se que elas

estejam envolvidas na estabilidade térmica em *archaeas* que enfrentam condições extremas, como em *P. furiosus*, por exemplo.



**Figura 4.** Organização das unidades estruturais já encontradas para o domínio periplasmático das OSTs espalhadas pelos três domínios.

Lizak e colaboradores (Lizak *et al.*, 2011) conseguiram obter a primeira estrutura completa de uma OST, sendo essa a PglB de *C. lari*. Dessa maneira, foi determinado que a PglB possui dois grandes domínios: um transmembrana e um periplasmático. (Figura 5) O domínio transmembrana compreende os resíduos 1-432, enquanto o periplasmático compreende do resíduo 433 ao 712. A porção transmembrana dessa proteína envolve 13 hélices inseridas na bicamada, além de duas alças externas (*external loop* – EL), chamadas EL1 e EL5. As hélices denominadas TM1 a TM4 e TM10 a TM13 promovem a região de interface com o domínio periplasmático, e formam a cavidade de interação com a proteína aceptora e também com o sítio catalítico da PglB. No outro lado, uma segunda cavidade é formada, onde provavelmente ocorrerá a interação com o oligossacarídeo ligado ao lipídio. O ponto de conexão entre as duas cavidades é encontrado justamente onde estará posicionada a cadeia lateral da asparagina aceptora e próximo ao sítio ativo da enzima. No que diz respeito ao domínio periplasmático, estão localizados dois motivos principais, bem caracterizados, envolvidos na acomodação e na interação firme da estrutura proteica para que ocorra a catálise. O primeiro deles é o motivo identificado como WWDYGY, primeiramente (Yan & Lennarz, 2002), e WWDXGX em um segundo momento, já que em algumas espécies ocorrem alterações nas tirosinas do motivo (Kamishnki *et al.*, 2013). O WWD inicial é indicado como sítio de reconhecimento do S/T da assinatura da N-glicosilação, formando ligações de hidrogênio entre esses resíduos. O outro motivo importante é o MI/DK, sendo que o MI e suas variações ocorrem em bactérias e em *archaeas*, enquanto o DK e suas variações ocorrem em *archaeas* e eucariotos. Dados cristalográficos (Lizak *et al.*, 2011) sugerem que a presença do motivo MI fornece uma explicação plausível para a preferência de uma treonina na posição +2 da assinatura, pois a cadeia lateral do resíduo de isoleucina poderia realizar interações hidrofóbicas com a metila presente na cadeia lateral de treoninas e ausente nas serinas.

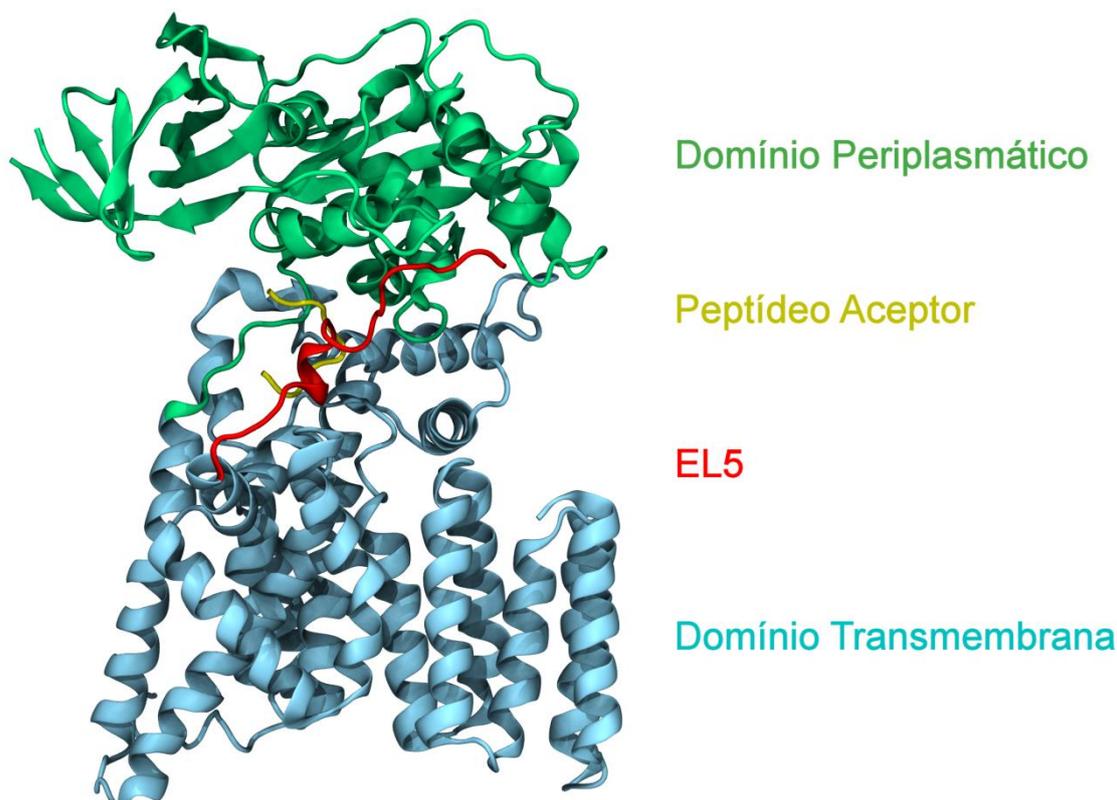
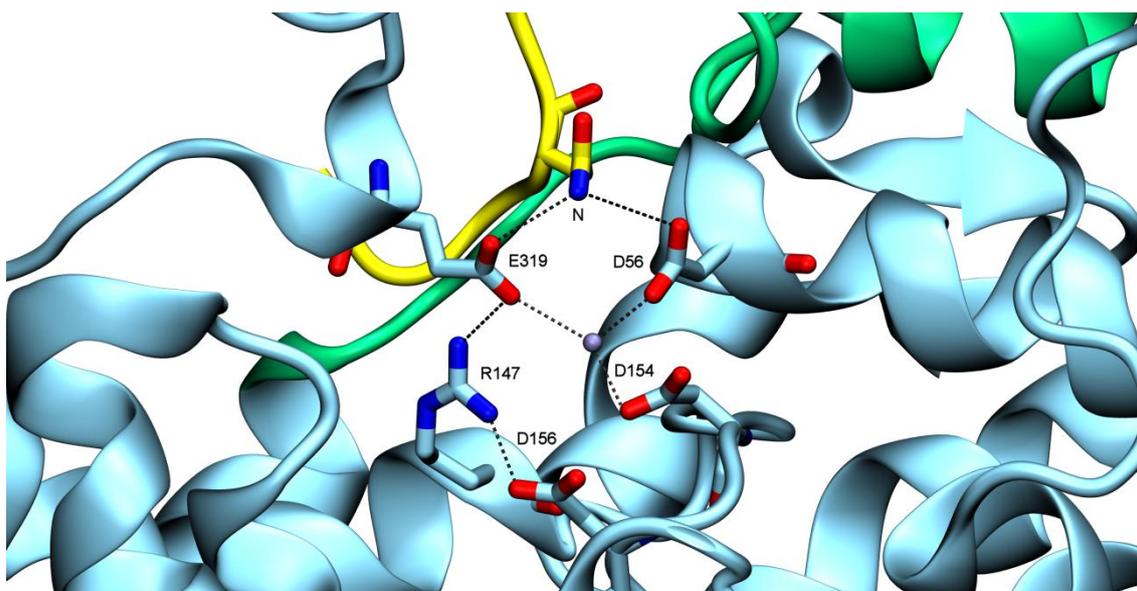


Figura 5. Estrutura da enzima PglB cristalizada juntamente com um peptídeo aceitor contendo a assinatura para N-glicosilação.

Com a elucidação da estrutura da PglB, cristalizada juntamente a um peptídeo aceitor com assinatura otimizada para maior eficiência na glicosilação, um mecanismo de ação para essa enzima pode ser proposto (Figura 6) (Lizak *et al.*, 2011). Para que o processo ocorra, é necessária a presença de um metal divalente ( $M^{2+}$ ), o qual se acredita que seja manganês ( $Mn^{2+}$ ). Na estrutura cristalizada, o metal era magnésio ( $Mg^{2+}$ ), todavia, a PglB também é ativa com esse íon, sendo essa uma propriedade conservada entre glicosiltransferases (GTases) (Lairson *et al.*, 2008). Os resíduos que pertencem ao sítio catalítico são dois aspartatos (Asp56 e Asp154) e um glutamato (Glu319), coordenados pela presença do metal  $Mg^{2+}$  (Figura 6). Ainda, esse metal aparenta ter uma função dupla, pois opera tanto na orientação dos aminoácidos do sítio como na estabilização do pirofosfato do lipídio carreador. Com a coordenação de um oxigênio de cada carboxilato da cadeia lateral do Asp56 e do Glu319, os outros dois oxigênios iriam formar ligações de hidrogênio com o grupo amida, presente na cadeia lateral da asparagina a ser glicosilada. Nesse ínterim, a partir da formação dessas interações, ocorreria a

rotação da ligação N-C do grupo amida, evitando a conjugação dos elétrons do nitrogênio com o grupo carbonila. Isto acarretaria em uma hibridização  $sp^3$  estaria presente no átomo de nitrogênio, bem como no posicionamento de seu par de elétrons que, assim, poderia realizar o ataque nucleofílico no carbono da posição 1 da Bac. Adicionalmente, a EL5 funcionaria como um elemento altamente flexível e desordenado na forma livre da enzima ao passo que, com a interação da proteína aceptora em sua respectiva cavidade, essa alça se acoplaria a mesma, mantendo assim o substrato posicionado e formando o sítio catalítico, visto que o Glu319 se localiza nessa região. Ainda, no que diz respeito à assinatura estendida para a N-glicosilação em bactérias, o resíduo Arg331 foi identificado como altamente conservado nesses organismos e, de fato, ele se posiciona em uma região próxima ao acceptor proteico, fornecendo possíveis contatos para o reconhecimento da assinatura e manutenção de uma interação forte com a OST.



**Figura 6.** Estrutura do sítio catalítico da PglB de *C. lari* (Lizak *et al.*, 2011). Uma rede de ligações de hidrogênios ocorreria juntamente à manutenção do íon divalente  $Mg^{2+}$  no centro, que por sua vez coordenaria tanto as cadeias laterais dos resíduos catalíticos como o lipídio carreador.

Um grande esforço tem sido feito na tentativa de inferir as relações evolutivas envolvendo as diferentes OSTs presentes nos três grandes domínios da vida, já que a resolução dessa questão poderia trazer benefícios na compreensão de mecanismos conservados entre essas moléculas. Todavia, ainda há uma lacuna nesse assunto, pois até o momento há apenas um estudo

prévio, sem suporte estatístico, envolvendo os três domínios (Maita *et al.*, 2010). Algumas das relações já foram descritas para algumas espécies de *archaeas* e bactérias, geralmente utilizando métodos como o *Neighbour-Joining* e não demonstrando todos os dados estatísticos em alguns casos (Tabish *et al.*, 2011; Kaminski *et al.*, 2013). Uma hipótese sugerida em estudos prévios indica que as bactérias teriam recebido a característica da N-glicosilação a partir do processo de transferência horizontal de genes, devido ao agrupamento de algumas sequências de *archaeas* da classe Methanobacteria, Methanococci e Thermoplasmata com as sequências de bactérias (Maita *et al.*, 2010). Nesse mesmo estudo, foi observado que essas sequências possuíam o mesmo motivo MIV ou suas variantes, dessa forma indicando que elas compartilhavam a ancestralidade. Contudo, uma das principais dificuldades para a realização de uma inferência filogenética adequada é a baixa identidade entre as sequências de OSTs, mesmo dentro de um mesmo domínio da vida apesar de, estruturalmente, elas possuírem arquiteturas similares.

#### 1.5 APLICAÇÕES BIOTECNOLÓGICAS E A GLICOENGENHARIA DE PROTEÍNAS

Um novo campo que começa a ganhar destaque dentre as potenciais aplicações biotecnológicas que a compreensão da via da N-glicosilação possibilita diz respeito à glicoengenharia de proteínas (Baker *et al.*, 2013). Atualmente, cerca de 70% das proteínas usadas na terapêutica, que estão aprovadas pelas agências regulatórias Europeia e Americana ou que estão em fase de desenvolvimento clínico ou pré-clínico, são glicosiladas (Sethuraman & Stadheim, 2006). Como já comentado, o sistema de glicosilação de *C. jejuni* já foi extensivamente caracterizado. Uma característica peculiar da PglB é a sua promiscuidade, demonstrada pela sua baixa especificidade de substrato. Estudos prévios (Wacker *et al.*, 2006; Chen *et al.*, 2007) indicam que o único requisito para uma glicana qualquer ser transferida por essa enzima é a presença de um grupo acetil na posição C-2 do monossacarídeo que será ligado à cadeia lateral da asparagina. Dessa forma, diversos experimentos apontam caminhos para o uso desse conhecimento no desenvolvimento de novas terapias.

Um dos primeiros trabalhos nesse sentido tratou da transfecção do locus gênico das enzimas da família Pgl de *C. jejuni* para *E. coli*, uma espécie que

não possui um sistema de N-glicosilação nativo (Wacker *et al.*, 2002). Posteriormente, diversos trabalhos descreveram a adição de diferentes tipos de estruturas sacarídicas, como os O-antígenos de espécies bacterianas, para auxiliar no desenvolvimento de vacinas (Feldman *et al.*, 2005; Ihssen *et al.*, 2010; Iwashkin *et al.*, 2012). Outras abordagens incluem a obtenção de tipos de cadeias oligossacarídicas presentes em eucariotos, como o núcleo inicial trimanosídico de mamíferos e o antígeno Lewis X (Valderrama-Rincon *et al.*, 2012; Hug *et al.*, 2011). Ainda, a glicosilação de outras biomoléculas também apresenta avanços. O lipídio A, um dos componentes do lipopolissacarídeo (LPS) que se faz presente na membrana de bactérias, é alvo de modificações para a formação de LPS quiméricos, contendo estruturas como as de gangliosídeos e antígenos Lewis (X e Y) (Focareta *et al.*, 2006; Paton *et al.*, 2005; Ilg *et al.*, 2010; Paton *et al.*, 2001; Mally *et al.*, 2013; Yavuz *et al.*, 2011). Na Figura 7 é mostrado um resumo da situação encontrada até o presente momento.

Como toda área em desenvolvimento, a glicoengenharia possui desafios que limitam a propagação de suas aplicações (Baker *et al.*, 2013). É necessária uma maior detecção e caracterização de novas estruturas oligossacarídicas, especialmente em bactérias e *archaeas* onde ocorre a maior variabilidade, bem como uma produção e conjugação mais eficiente de glicanas eucarióticas em bactérias, buscando a expansão das ferramentas disponíveis. Também importante é encontrar um melhor entendimento das vias de N-glicosilação, visto que elas possuem variações interespecíficas, propiciando assim uma melhor manipulação desses sistemas para o crescimento do campo. Por fim, o aperfeiçoamento da compreensão e da predição das consequências no enovelamento e na função de proteínas, acarretadas pela glicosilação em sítios específicos dessas, possibilitará avanços no desenvolvimento e no desenho específico de glicanas com diferentes utilidades.

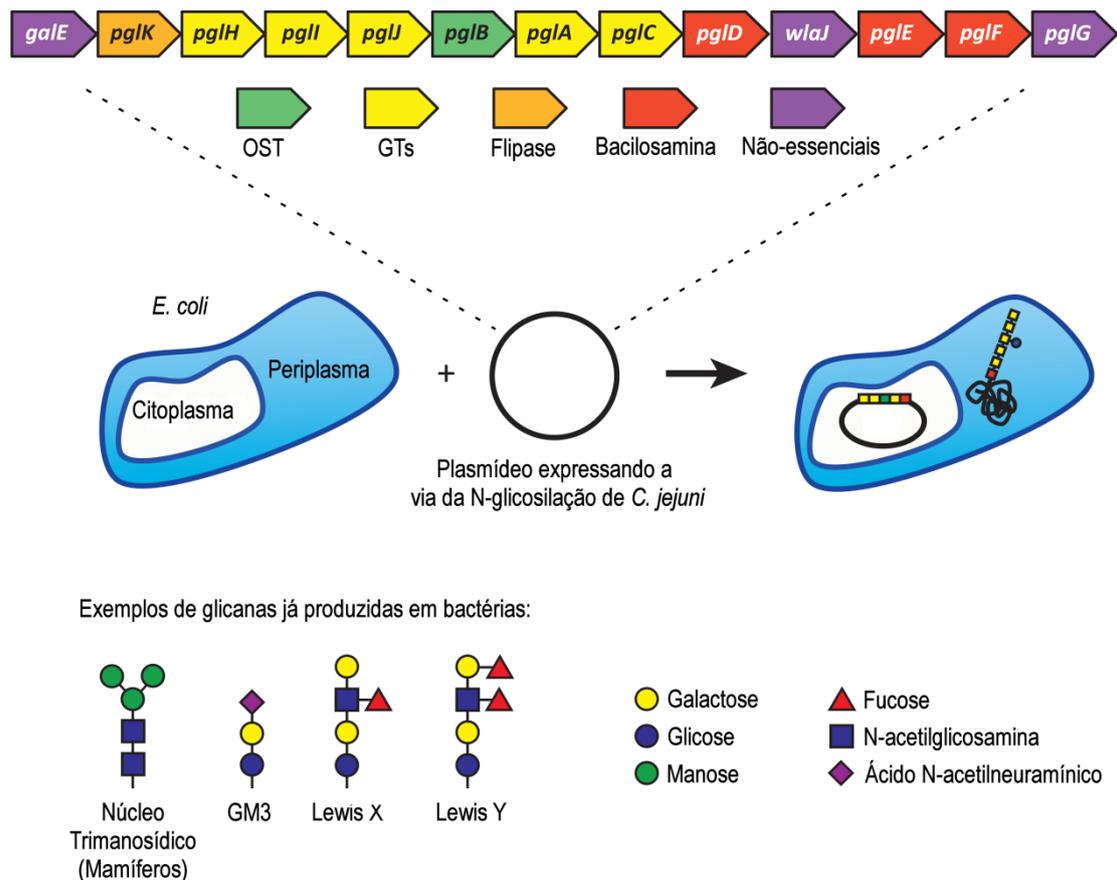


Figura 7. Representação esquemática da engenharia envolvida na expressão de N-glicosilação em *E. coli*. (adaptado de Baker *et al.*, 2013).

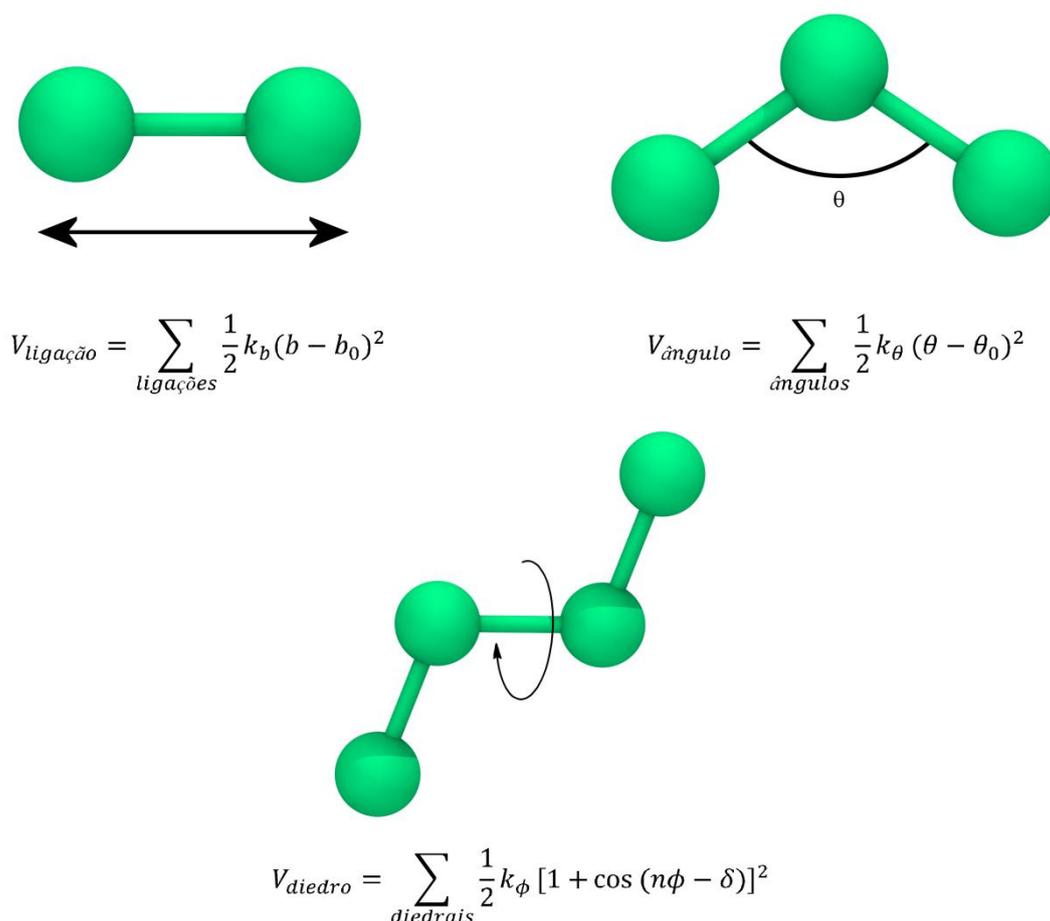
## 1.6 DINÂMICA MOLECULAR

A técnica de dinâmica molecular (DM) é uma abordagem que vem sendo empregada no estudo de biomoléculas das mais variadas, servindo de auxílio para a compreensão e interpretação de dados experimentais obtidos tanto em nível atômico como molecular. É usualmente aplicada em comparações com dados obtidos via cristalografia de raios-X e ressonância magnética nuclear (RMN) (van Gunsteren *et al.*, 2006). Consiste em uma simulação computacional, onde átomos e moléculas podem interagir por um determinado período de tempo e serão ajustados a partir de parâmetros físico-químicos envolvendo os componentes desse sistema. Sua base refere-se à mecânica clássica (Leach, 2001), sendo ela regida pela resolução da segunda lei de Newton:

$$F_i = m_i \cdot a_i$$

, onde uma determinada força  $F_i$  é aplicada a um átomo  $i$  com massa  $m_i$ , gerando uma aceleração  $a_i$ . Essa equação é sucessivamente integrada e é resolvida para todos os átomos do sistema, acarretando em uma avaliação do movimento ou trajetória, tanto global como local, de um determinado sistema, em função do tempo ( $\Delta t$ ). A informação é armazenada em forma de trajetória que pode variar em tamanho, dependendo do sistema, pelo tempo de integração definido, sendo este, em geral, de um ou dois femtossegundo(s).

O chamado Campo de Força é o conjunto de potenciais que serão responsáveis por descrever as interações dos átomos durante a DM, permitindo tanto a análise de interações ligadas e não-ligadas. Esses potenciais são aplicados na descrição das ligações entre dois átomos, bem como de ângulos entre três átomos e ângulos torsionais (diedros) compreendendo quatro átomos, exemplificados abaixo:



**Figura 8.** Potenciais envolvidos na descrição dos termos do campo de força que será empregado em simulações por dinâmica molecular.

Outras interações também compõem o conjunto de forças que participam da simulação por DM. As interações não-covalentes, como interações de van der Waals, podem ser calculadas com o potencial de Lennard-Jones, assim como as interações eletrostáticas.

## 2. OBJETIVOS

Considerando que as variadas formas encontradas de OSTs na natureza ainda carecem de informações estruturais adicionais, especialmente devido à dificuldade na obtenção do domínio transmembrana, e o estudo desses sistemas pode proporcionar a extrapolação dos mecanismos encontrados na PglB bacteriana para outras enzimas do mesmo tipo, o presente trabalho busca contribuir no entendimento da glicobiologia estrutural dessa proteína. Duas áreas principais serão enfocadas: a) a caracterização da estrutura e da dinâmica da PglB em membrana e solução; b) a análise filogenética da ancestralidade da PglB e suas homólogas em outros domínios.

Os seguintes objetivos específicos foram traçados para o desenvolvimento do trabalho:

- Observar a dinâmica dos resíduos catalíticos e a sua influência pelo íon  $Mg^{2+}$ ;
- Avaliar o comportamento de aminoácidos considerados importantes tanto para o reconhecimento dos substratos como para a catálise propriamente dita;
- Analisar as alterações estruturais geradas em decorrência da presença ou ausência de substratos componentes do sistema;
- Inferir as relações de ancestralidade e possíveis variações em motivos conservados do domínio Bacteria;
- Realizar a caracterização filogenética, com suporte estatístico compreendendo organismos dos três domínios da vida.
- Buscar novas correlações entre regiões conservadas no alinhamento com aspectos estruturais observados a partir das simulações e relacionar aos clados encontrados nas análises filogenéticas.

### 3. MÉTODOS

#### 3.1 PROGRAMAS UTILIZADOS

Os principais programas e protocolos e em quais partes da presente dissertação eles foram empregados encontram-se abaixo:

Tabela 1. Descrição dos programas e/ou parâmetros empregados para a realização de diversas tarefas.

Finalidade	Programas e/ou Parâmetros
Visualização e Manipulação de Estruturas	VMD 1.9.1 (Humphrey <i>et al.</i> , 1996), PyMol (DeLano, 2002) e MOLDEN (Schaftenaar & Noordik, 2000)
Modelagem de Alças	FALC-Loop (Lee <i>et al.</i> , 2011)
Análise Estereoquímica	PROCHECK (Laskowski <i>et al.</i> , 1993)
Inserção em Membrana	InflateGRO (Kandt <i>et al.</i> , 2007)
Dinâmica Molecular	GROMACS 4.5.1 (Pronk <i>et al.</i> , 2013), GROMOS 53a6 (Oostenbrink <i>et al.</i> , 2004) e GROMOS 53A6 <sub>Glyc</sub> (Pol-Fachin <i>et al.</i> , 2012)
Geração de Topologias	PRODRG Beta v2.5 (Schuettelkopf & van Aalten, 2004)
Inferência Filogenética	SimPlot 3.5.1 (Lole <i>et al.</i> , 1999), FigTree v1.4, MEGA 5.1 (Tamura <i>et al.</i> , 2011) e MrBayes (Huelsenbeck <i>et al.</i> , 2001)

#### 3.2 PREPARAÇÃO DO SISTEMA

A PglB possui sua estrutura cristalizada, depositada no banco de dados do PDB, sob código 3RCE (Lizak *et al.*, 2011). Essa estrutura cristalizada possui duas regiões onde faltam aminoácidos em relação a sua sequência original: a primeira delas é localizada no domínio transmembrana, mais especificamente no EL5, compreendendo os resíduos 283 a 306; a segunda é

localizada no domínio periplasmático, na região contendo o barril- $\beta$  da unidade IS, compreendendo apenas três resíduos do 605 ao 607. Essas regiões foram preenchidas a partir da submissão dessa estrutura ao servidor FALC-Loop (Ko *et al.*, 2011), que utiliza uma metodologia (Lee *et al.*, 2010) envolvendo construção por fragmentos coletados em bancos de dados de estrutura e utilização do algoritmo *analytical loop closure* em conjunto com minimizações de energia que assim satisfazem ângulos de torção apropriados para os resíduos da alça modelada. Múltiplas estruturas são fornecidas pelo servidor, possibilitando a escolha do modelo mais apropriado para cada situação. Como a região do EL5 possui uma grande quantidade de aminoácidos faltando e essa é considerada importante para o mecanismo da catálise, a seleção do melhor modelo foi feita visando essas características. Assim, o modelo selecionado foi aquele que possuiu melhor escore na avaliação da estereoquímica e que manteve a abertura da suposta cavidade de ligação do substrato doador aberta. A partir da estrutura escolhida, foram construídos três sistemas inseridos em membrana: i) PglB na forma livre, apenas com a presença do íon coordenando o sítio catalítico; ii) PglB complexada com um peptídeo de assinatura ótima para N-glicosilação (presente no próprio cristal da PglB; iii) PglB complexada ao peptídeo glicosilado com o heptassacarídeo de *C.jejuni*.

### 3.3 INSERÇÃO EM MEMBRANA

Buscando melhor emular a condição em que se encontra nos organismos, a PglB foi inserida em uma membrana de 1-palmitoil-2-oleil-sn-glicero-3-fosfoetanolamina (POPE) com 512 fosfolipídios. A metodologia utilizada foi a inflateGRO (Kandt *et al.*, 2007). Ela consiste em sobrepor a estrutura da PglB em membrana na posição onde deseja-se inseri-la para, subsequentemente, expandir as dimensões da caixa de simulação e mover os lipídios lateralmente, multiplicando essas dimensões por um fator de escalonamento comum. Após esse processo, os lipídios serão lentamente comprimidos em direção ao centro da caixa, juntamente com as dimensões dessa, enquanto uma série de minimizações é realizada durante esse retorno, até que o sistema atinja a área por lipídio inicial apropriada. A proteína permanece com uma força de restrição de posição, evitando que sua estrutura seja afetada. Qualquer lipídio que se mantenha em contato com a proteína

após a fase de expansão da caixa será removido. Portanto, os sistemas finais foram simulados com membranas possuindo 490 moléculas de POPE.

### 3.4 DINÂMICA MOLECULAR

Para o estudo por DM, o pacote de simulações do GROMACS 4.5.1 (Pronk *et al.*, 2013) foi utilizado, em conjunto com os campos de força GROMOS 53A6 (Oostenbrink *et al.*, 2004) adicionado dos parâmetros GROMOS 53A6<sub>GLYC</sub> para a parte sacarídica (Pol-Fachin *et al.*, 2012 e resultados não publicados). Os sistemas foram solvatados a partir do modelo de água SPC (Berendsen *et al.*, 1987) em caixas retangulares com a presença de condições periódicas de contorno. As minimizações de energia foram conduzidas no algoritmo *Steepest Descent*. Além disso, para o tratamento eletrostático, o método PME (Darden *et al.*, 1993) foi empregado. Os comprimentos das ligações covalentes foram fixados pelo método de Lincs (Hess *et al.*, 1997), permitindo um tempo de integração de 2 fs para as simulações subsequentes.

As simulações por DM podem ser divididas em duas fases: a fase de relaxamento e a fase de produção. No relaxamento, a biomolécula de estudo terá uma restrição de posição imposta, visando à obtenção de uma acomodação lenta e gradual da orientação dos outros componentes do sistema, como, por exemplo, solvente e contraíons. Assim, duas simulações, uma de 0,0001  $\mu$ s e outra de 0,001  $\mu$ s, foram realizadas para cada sistema, com uma constante de força 1000 kJ/mol. A primeira dessas foi realizada com volume e temperatura constantes (NVT) para que a temperatura seja adequadamente ajustada, enquanto as posteriores foram realizadas com pressão e temperatura constantes (NPT) para que a pressão e a densidade correta da caixa sejam alcançadas. Na fase de produção (também NPT), onde ocorre a coleta dos resultados, todo o sistema (PglB e membrana) tem seu movimento liberado da penalidade imposta, possibilitando a avaliação da dinâmica. Nesta fase, o tempo de simulação foi de 0,2  $\mu$ s, objetivando a acomodação da membrana e da proteína. Na simulação de NVT, o acoplamento de temperatura foi realizado com o termostato V-rescale (Bussi *et al.*, 2007). Nas simulações de NPT, como essas envolvem uma membrana

biológica, foi empregado o termostato Nosé-Hoover (Nose, 1984; Hoover, 1985). Quanto ao acoplamento de pressão, o barostato Parrinello-Rahman (Parrinello & Rahman, 1981; Nose & Klein, 1983) foi utilizado.

### 3.5 INFERÊNCIA FILOGENÉTICA

A caracterização das relações evolutivas da proteína PglB iniciou-se a partir da busca de sequências de aminoácidos presentes no banco de dados do *National Center for Biotechnology Information* (Sayers *et al.*, 2012). Duas abordagens foram utilizadas: a primeira foi a busca por similaridade de sequências, a partir do servidor *Basic Local Alignment Search Tool* (BLAST) (Altschul *et al.*, 1990), com a sequência de *C. lari* como entrada; a segunda foi uma busca manual utilizando as palavras-chaves “PglB”, “AglB” e “Stt3”, realizando uma filtragem por sequências que possuíam o motivo conservado WWDYGY. Após essa varredura, um total de 355 sequências foram obtidas para os três grandes domínios, sendo 49 de bactérias, 143 de *archaeas* e 163 de eucariotos. Para o alinhamento das sequências, o algoritmo ClustalW (Larkin *et al.*, 2007) foi empregado.

Buscando avaliar as possíveis regiões hipervariáveis, o programa SimPlot foi utilizado. Os dados foram divididos em quatro subgrupos: Archaea, Bacteria, Eukarya e o gênero *Campylobacter*. A partir do software MEGA5.1, o melhor modelo de substituição foi definido como o de Whelan and Goldman (WAG), além de ser incorporada a frequência média de aminoácidos dos dados utilizados (+F), juntamente com a distribuição Gamma para lidar com as diferentes taxas de evolução de cada sítio (cinco categorias, +G), somado ainda a prevalência de alguns sítios invariantes evolutivamente (+I) (Whelan & Goldman, 2001). Com o mesmo *software*, árvores filogenéticas foram calculadas pelo método de Máxima Verossimilhança (MV) e Evolução Mínima (EM), aplicando um valor de *bootstrap* (replicatas) de 1.000. Finalmente, foi também utilizado o software MrBayes (Huelsenbeck and Ronquist, 2001) para calcular árvores por Inferência Bayesiana (IB) durante  $6 \times 10^6$  gerações, empregando uma frequência de amostragem de 10. O software FigTree 1.4 (Rambaut, 2012) foi utilizado na manipulação e edição das árvores.

## 4. RESULTADOS

Os resultados dessa dissertação serão divididos em dois itens, 4.1 e 4.2. O primeiro contém os resultados do trabalho focado na temática da estrutura e dinâmica da PglB, enquanto o segundo irá ter como base a análise filogenética que está sendo realizada, entretanto ainda como resultados preliminares.

4. 1 STRUCTURAL AND DYNAMICS ASSESSMENT OF THE BACTERIAL  
OLIGOSACCHARYLTRANSFERASE PGLB

# Structural and Dynamics Assessment of the Bacterial Oligosaccharyltransferase PglB

Conrado Pedebos and Hugo Verli\*.

Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP  
15005, Porto Alegre 91500-970, RS, Brazil

\*Corresponding Author. TEL.: +55 51 3308 7770; fax: +55 51 3308 7309. Email address:  
[hverli@cbiot.ufrgs.br](mailto:hverli@cbiot.ufrgs.br) (H. Verli)

## Abstract

N-glycosylation is a widespread posttranslational modification, already identified in the three domains of life. Since its discovery in Bacteria, many studies have sought to evaluate the biochemical pathway and to unravel structural information of its components. The last step of this process involves an oligosaccharyltransferase (OST), which has homologues in all domains and is responsible for the reaction catalysis of the *en bloc* transfer of a fully assembled oligosaccharide chain to a protein possessing the extended N-linked glycosylation motif D/E-X-N-X-S/T. In Bacteria, this enzyme is the PglB, a protein of 712 residues and molecular weight of 84 kDa. Recently, this molecule had its full structure elucidated, enabling the description of its transmembrane and periplasmic domains, as well as the proposition of a catalytic mechanism involving three residues that are located at the transmembrane section. In the current paper, we evaluate the structural aspects of PglB by performing molecular dynamics simulations on three different systems, each corresponding to a unique state of this enzyme: unbound PglB, in its native state; PglB with the co-crystallized peptide acceptor; and, PglB with a glycosylated peptide. All systems were embedded in a palmitoyl-oleoyl-phosphatidylethanolamine (POPE) membrane model, further increasing the similarities with the protein environment. By performing this analysis, we were able to suggest prevalent motions that may bring insights into the mechanism of the glycosylated peptide detachment. Moreover, transient states related to the catalytic site of PglB could be observed, agreeing with previously proposed hypotheses of hydrogen bonds networking in that area, as well as demonstrating the flexibility of the acceptor Asn side chain.

Keywords: N-glycosylation; Molecular Dynamics; OST; PglB; Computational Biology.

N-glycosylation is among the most prevalent protein modifications in nature.<sup>1</sup> This process is found widespread in all domains of life and is mainly associated to signaling, stability, trafficking, adhesion, intracellular location, immune response in Eukaryotes,<sup>1</sup> protection against the harsh environment and flagella assembly in Archaea,<sup>2,3</sup> and host adhesion, colonization and invasion in Bacteria.<sup>4,5</sup> In the latter, this posttranslational changes was not reported until 1999.<sup>6</sup> In the N-glycosylation pathway, after the full assembly of the oligosaccharide chain, the responsible for the catalysis of the *en bloc* transfer is the oligosaccharyltransferase (OTase) enzyme PglB.<sup>7</sup> Recently, this biomolecule had its full structure characterized by studies of X-ray crystallography,<sup>8</sup> allowing a more comprehensive look at the complex array of interactions behind the mechanism of action concerning the oligosaccharide transfer. PglB structure consists of a transmembrane (TM) domain, from residue 1 to 432, and a periplasmic domain (PP), from residues 433 to 712. The transmembrane portion comprises 13 helix inserted in the bilayer and two external loops named EL1 and EL5. The EL5 structure is believed to play an important role in catalysis, since one of its residues, the Glu319, is a part of the catalytic site. The other two residues that compose the catalytic site are the D56 and D154, both presented in the TM domain. Besides, a conserved residue, R331, was detected as providing interactions with the extended glycosylation sequon (D/E in position -2).<sup>8</sup> As for the periplasmic domain, a classification of structural units was proposed.<sup>9,10</sup> Therefore, PglB exhibits two different units, the central core (CC) and the insertion sequence (IS) units. The CC unit presents the largely described conserved motifs WWDYGY and MIV. The WWD and the I residues from this motifs are specifically involved in the firm binding of the +2 position residue of the N-glycosylation sequon (S/T). Furthermore, the Y468 from the first motif has been cogitated as responsible for carbohydrate binding at the C2 substituent

position,<sup>7</sup> after a tentative modeling of this substrate on its cavity. Finally, another motif has been identified recently, the DxxK, which is also implicated as associating with the glycan chain.<sup>11</sup> The other structural feature, the IS unit, has a  $\beta$ -barrel-like structure containing 9 beta-strands in a  $\beta$ -sandwich structure. Although its function is not yet determined, it is hypothesized that it is dispensable for catalysis<sup>12</sup> and may be important for structural stability.<sup>10</sup>

Accounting for the mechanism of action of the enzyme, two cavities are proposed: one being the acceptor protein binding site, requiring not only the consensus sequence, but also a largely surface-exposed and flexible loop display,<sup>13</sup> while the other is a hydrophobic groove where the lipid portion of the LLO would be positioned,<sup>8</sup> allowing the interaction of both substrates through a communication area between cavities. The residues from the catalytic site would participate in a hydrogen bond network along with the protein substrate and the divalent ion  $Mg^{2+}$ , which results in an optimal position and configuration of the nitrogen atom from the amide of the Asn side chain to perform a nucleophilic attack on the C1 position from the Bacilosamine sugar. However, since the crystal resolution available is of 3.4 Å, such coordination of interactions remains not fully understood, as well as the correct position of the amide nitrogen and the carbonyl oxygen.<sup>11</sup>

In the last few years, PglB had many evidence collected regarding the biochemical and structural features of its catalytic motifs and other important regions, mainly employing mutagenesis<sup>14</sup> and topology approaches.<sup>11</sup> Many of them are conserved throughout the three domains of life, an element that highlights the importance of this enzyme as a model for the process of N-glycosylation. Therefore, intending to contribute with novel insights to the understanding of the glycobiology of this OST, the present work aims to describe the dynamics of PglB on its unbound,

native state, as well as with the protein substrate both non-glycosylated and glycosylated, applying computational techniques, such as molecular modeling, molecular dynamics (MD) simulations and metadynamics. Specifically, we attempt to further analyze and compare the different implications of ligand binding to the structure of PglB, as well as describe the conformational rearrangements occurring in a time dependent manner.

## **Results and Discussion**

Departing from the crystal structure with the missing residues already modelled, we performed 200ns MD simulations and were able to analyze the structural features comprising PglB structure. Initially, the root mean square deviation (RMSD) relative to the starting input structure of each run was evaluated (Figure 1) as a matter of the full structure, as well as separated by the functional domains. Also, the same analysis was applied for the acceptor peptide, already present at the original crystal structure. Despite the crystal resolution of 3.4 Å, the RMS remained reasonably stable throughout the simulations, showing lower values as the complexity of the observed systems increased. The higher RMS value corresponded to the PglB on its native state (black), which was expected since the original crystal structure was obtained, co-crystallized with its protein substrate. By analyzing the contribution of the enzyme domains for the total RMS, we observed that a lower transmembrane value contributes to the lower total value of the PglB complexed with the glycopeptide (green), while the periplasmic domain presented no detectable difference between all runs. Additionally, concerning both peptides measures, no significant deviations were also found between them, since the peptide appears to be firmly attached to the binding cavity throughout the simulation.

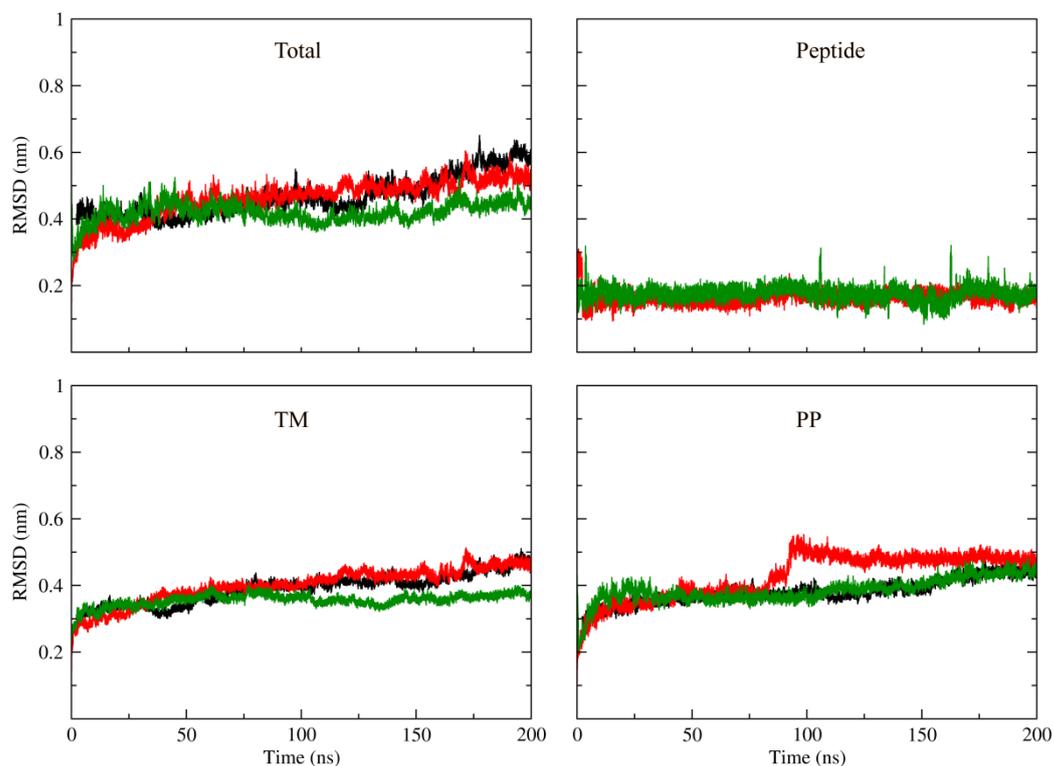


Figure 1. RMSD analysis comprising the 200ns MD simulations of the PglB at the native state (black) and complexed with the non-glycosylated peptide (red) and with the glycosylated peptide (green). Results are shown for the total protein, for the bound peptide, for the transmembrane domain (TM) and for the periplasmic domain (PP).

Further depicting the structural plasticity of PglB, several regions demonstrated variations due to the presence of the substrates in the cavities, as can be seen in the root mean square fluctuation (RMSf) extracted from the principal component analysis (PCA). (Figure 2) The presence of both non- and glycosylated peptide affects the behavior of three loops comprising residues 68 to 78, 291 to 304 (initial portion of EL5) and 516 to 526, which are located circumventing the glycan cavity site. Fitting frames of the three simulations at every 20 ns, we observed the movement of these loops,

further generating the closure of the cavity in both complexed states; while in the unbound state the same did not occur. This may be explained by the presence of the peptide which binds to the acceptor cavity and the EL5, reducing the motility of the OST and thus promoting an increase in the interactions between the exposed structures.

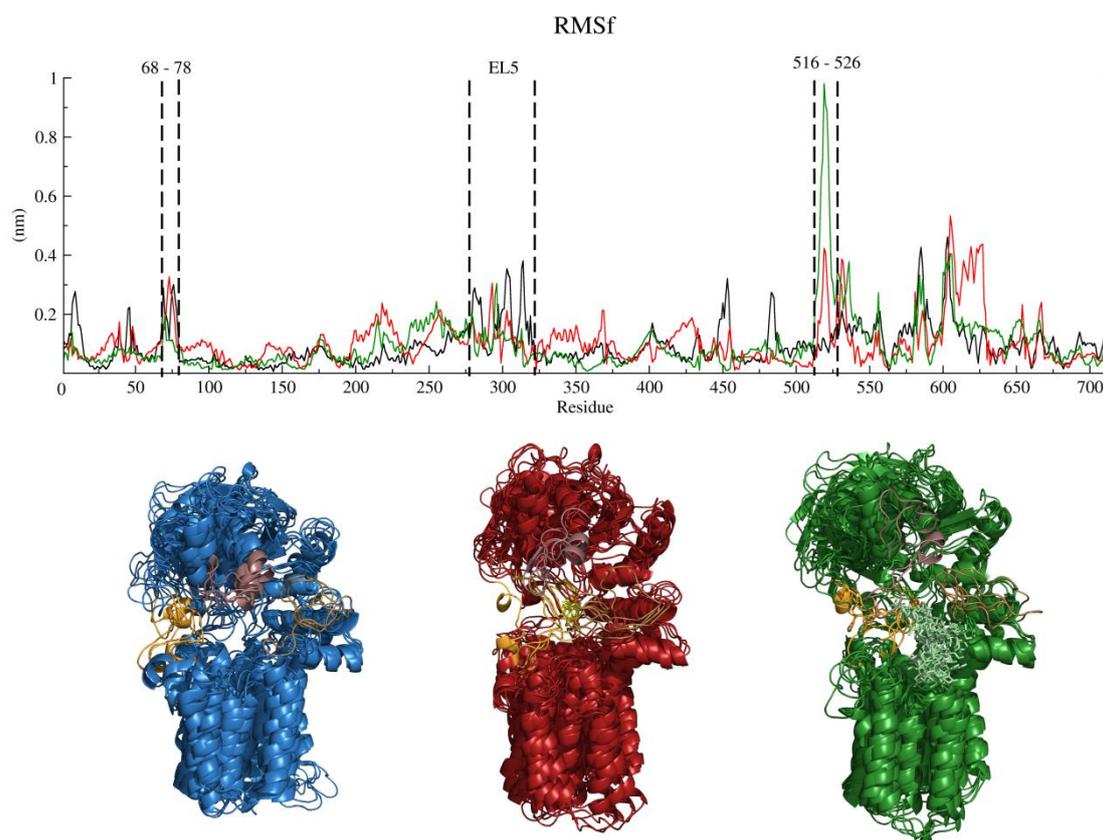


Figure 2. A) RMSf analysis for the PglB at the native state (black) and for the complexed form with the non-glycosylated peptide (red) and with the glycosylated peptide (green). B) Superimposition of frames at each 50ns for PglB at the native state (blue) and for the complexed form with the non-glycosylated peptide (red) and with the glycosylated peptide (green). The brown area represents the 516-526 portion, while the orange represents the 68-78 portion and the yellow represents the EL5 section.

The large amount of interactions induced by the glycosylated peptide agrees with the steric tension hypothesis proposed by Lizak *et al.*,<sup>8</sup> although the ligand never exited the protein during our simulations. Further investigating this phenomenon, we calculated the linear interaction energy (LIE) of both ligands so that we could estimate the free energy of binding ( $\Delta G_{\text{binding}}$ ) of each substrate. The obtained value for the non-glycosylated peptide was -16.42 kJ/mol, while the value for the glycosylated peptide was -64.25 kJ/mol. However, considering the large standard deviations of  $\pm 48.31$  and  $\pm 52.02$ , respectively, the binding energies are overlapped, indicating similar affinities for both ligands, with a slight preference for the glycosylated peptide. Therefore, this data suggests that another kind of mechanism might be behind the glycoprotein moving out, since we expected a higher affinity for the non-glycosylated peptide. Examining the structures comprehended in between the extreme projections of the first principal component (eigenvector 1), which is the most representative, it is possible to observe the transition between two conformational states (Figure 3). As shown, the periplasmic portion performs a twist motion, where the regions circumventing the glycan cavity shift their structures towards the carbohydrate portion of the glycopeptide. Concomitantly, the  $\beta$ -barrel-like IS unit appears to have an upward motion, allowing the expansion of the protein binding cavity for the glycopeptide departure. Therefore, we propose that an allosteric regulation may occur between sites, facilitating the entrance and the exit of the enzyme substrates.

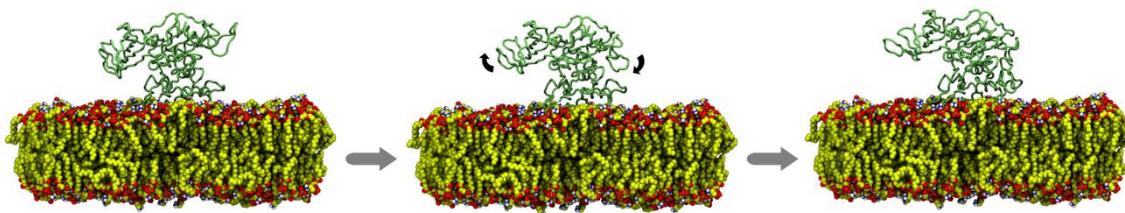


Figure 3. Schematics of the structures extracted from the extreme points of the first principal component, depicting the conformational transition observed for both cavities.

The three residues from the catalytic site of PglB are predicted to interact with a catalytic divalent metal ion ( $M^{2+}$ ), which performs the coordination of these residues. Consequently, we measured the distances between these atoms and the co-crystallized  $Mg^{2+}$ , aiming to evaluate if they were compatible with metal coordination (Table 1).

**Table 1.** Distances between the catalytic site ion  $Mg^{2+}$  and the side-chains from the catalytic residues for each analyzed system.

Residue and atom	Distance to $Mg^{2+}$ (Å)		
	PglB Native	PglB + non-glycosylated peptide	PglB + glycosylated peptide
Asp56 O $\delta$ 1	2.14±0.08	2.09±0.01	2.14±0.10
Asp56 O $\delta$ 2	2.07±0.06	2.12±0.08	2.12±0.08
Asp154 O $\delta$ 1	2.08±0.07	2.07±0.06	2.10±0.07
Asp154 O $\delta$ 2	2.08±0.06	2.14±0.31	2.08±0.06
Glu319 O $\epsilon$ 1	2.10±0.07	2.14±0.08	2.08±0.06
Glu319 O $\epsilon$ 2	2.08±0.07	2.10±0.07	2.13±0.08

Looking at the distances calculated in our models, the  $Mg^{2+}$  ion was firmly stable between residues in the studied simulations, as a result of interaction with the six oxygen atoms presented, where each residue can contribute with either a monodentate and/or a bidentate coordination. Such observed distances are compatible with the octahedral geometry, which is the most prevalent in ions  $Mg^{2+}$ , and are in closer agreement with experimental data (2.073) than the crystal structure.<sup>15</sup>

The previously proposed catalytic mechanism by observations of the crystal structure postulates that the Asp56 and the Glu319 would form hydrogen bonds with the amide of the acceptor Asn. This trait would permit an increase in the electronegativity of the poor nucleophile that amides usually are, modifying its geometry from the planar

conformation to the  $270^\circ$  conformation, which exposes a lone pair of electrons to perform a nucleophilic attack. Hence, we calculated the distances between the oxygens from the  $\text{COO}^-$  groups of both residues side chains and the hydrogen atoms of the amide (Figure 4).

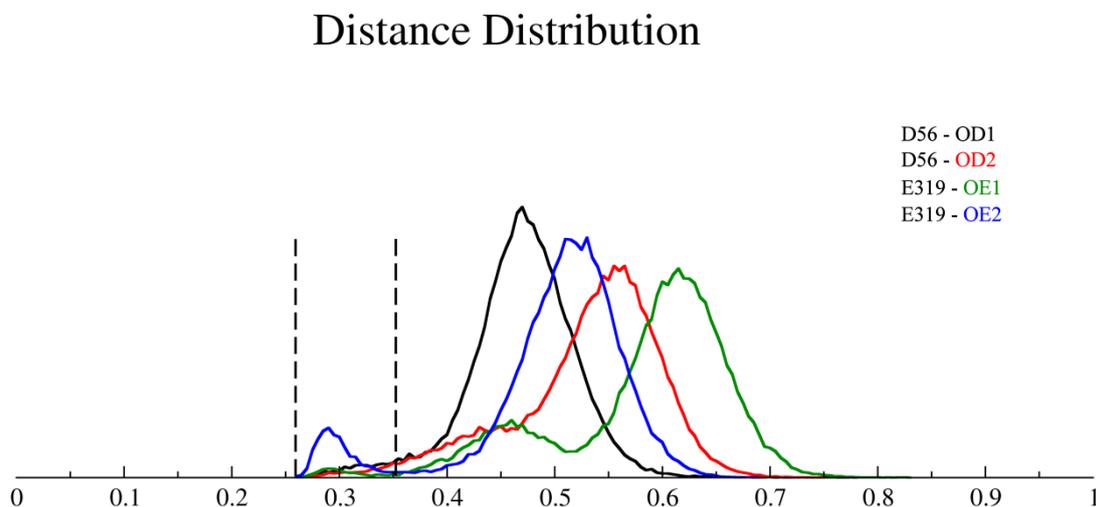


Figure 4. Distance distribution between the catalytic residues side-chain and the acceptor amide from the Asn side-chain, highlighting the minor state where distances are appropriate for hydrogen bonding.

The distances between the  $\text{N}\delta 2$  of the acceptor Asn and the Asp56 two oxygen  $\text{O}\delta 1$  and  $\text{O}\delta 2$  are, respectively,  $4.70 \pm 0.50$  and  $5.30 \pm 0.68$ , while for the Glu319 two oxygen  $\text{O}\epsilon 1$  and  $\text{O}\epsilon 2$ , the distances are  $5.81 \pm 0.83$  and  $4.99 \pm 0.71$ . These averages, nonetheless, do not reflect the whole behavior of these atoms. Observing the distribution of these values along time (Figure 4), we identify two population peaks, one major and one minor. The minor peak indicates a presence of the distances close enough to account for a hydrogen bond. Thus, we calculated hydrogen bonds by making use of the gromacs tool *g\_hbond*, assessing the occurrence of these interactions during the simulation (Figure 5).

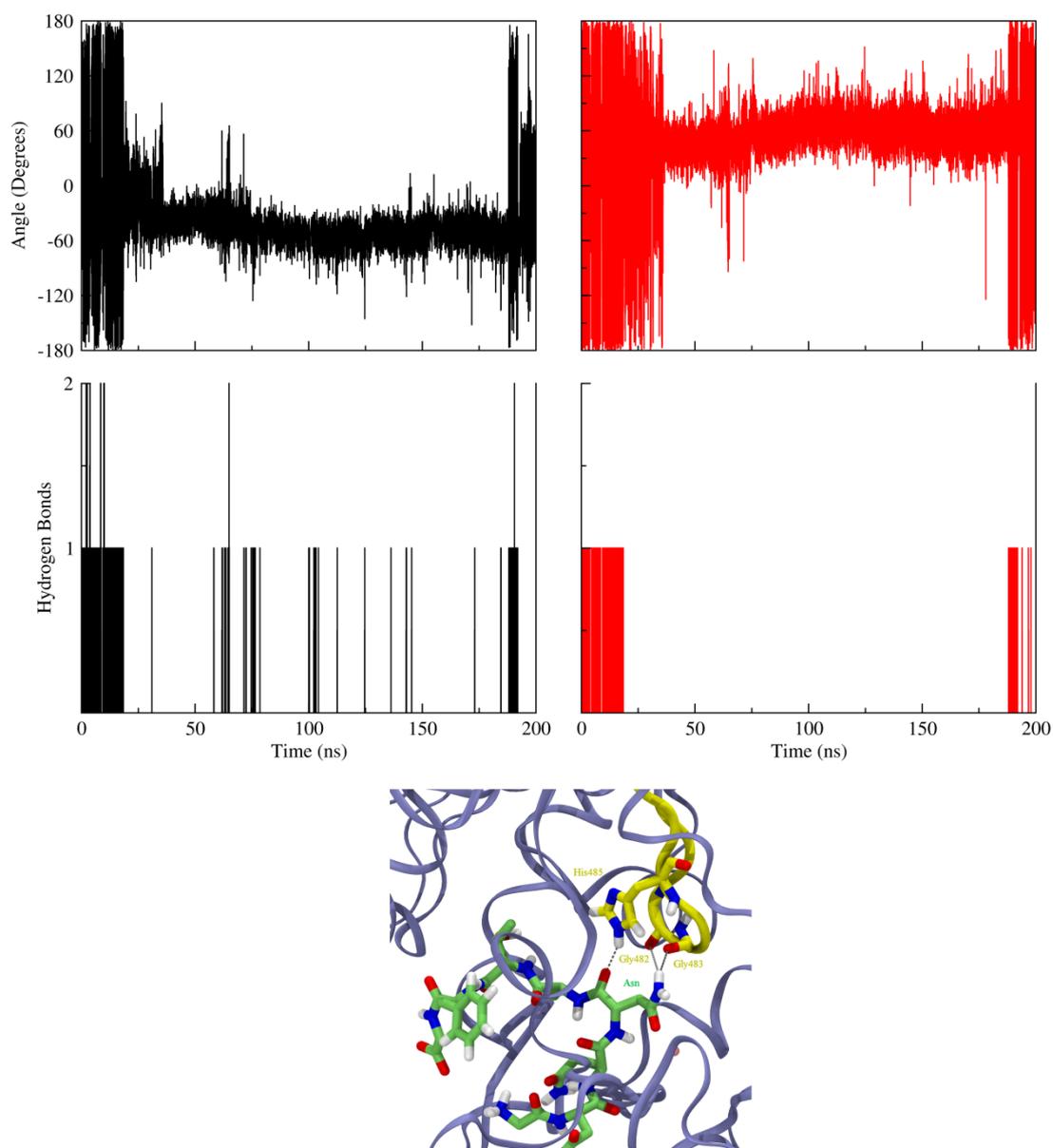


Figure 5. A) Reorientation of the amide and carbonyl dihedrals during the 200ns simulation demonstrating the high fluctuations at the beginning and at the end, while the rest remains stable. B) Hydrogen bonding accounting for the side-chains of the catalytic residues, correlating with the dihedral transitions. C) Interactions between the DxxK motif with the acceptor Asn side-chain and backbone.

Apparently, most of the contacts occur in the beginning of the simulation and in the end, with a more prominent role of the residue Asp56. Indeed, around the first 20 ns of the simulation, a reorientation of both the amide and the carbonyl dihedral angles ( $C\beta-C\gamma-N\delta 2-Mg$  and  $C\beta-C\gamma-O\delta 1-Mg$ ) from the Asn-acceptor side chain was detected (Figure 5), causing the amide to drift apart from the  $Mg^{2+}$  ion and the catalytic residues, interacting with the loop comprehending Asp481 to Gly487, specially with Gly482, Gly483 and His484 (Figure 5). In addition, the carbonyl from the Asn side chain stays at coordination distance from the catalytic ion, with a 4.08 Å in average, from time to time.

This scenario remains at most of the time of the simulation, until the last 20 ns, when another rearrangement of this side-chain occurs, returning to a state similar to the beginning of the simulation. Employing a radial distribution function (RDF) analysis (Figure 6A) to verify the presence of solvent atoms near the  $Mg^{2+}$  ion, we detected the presence of at least two hydration layers. By taking snapshots of the simulation, it became possible to recognize three patterns of water metal coordination: the first one (Figure 6B) occurs before the inversion of the amide nitrogen, while the second one (Figure 6C) is after the inversion and happens with the carbonyl coordinating together; the third one (Figure 6D) is also after the inversion, however the water molecule now occupies the position where the carbonyl was previously and, in addition, coordinates this group. Also, in the second pattern, another water molecule is interacting with the COO- group of Glu319. A shared commonality of the three configurations is that the octahedral coordination is maintained. All these data suggests that not only the coordination presented in the crystal structure is prevalent, but that are also at least a few variations in the catalytic site and in the position of the substrate. Moreover, as previously discussed,<sup>11</sup> the crystal resolution of 3.4 Å, together with the high pH of

crystallization (9.4), may interfere with the proposed mechanism involving the network of hydrogen bondings between the catalytic residues and the acceptor Asn. What we observe is a dual behavior, where not only hydrogen bonds are formed in some time windows of the simulation, but also the Asn side chain demonstrates a flexibility of position and interaction partners. Besides, the interaction of this residue with the DxxK motif may act as another mechanism that keeps the amide group positioned near the substrate donor site.

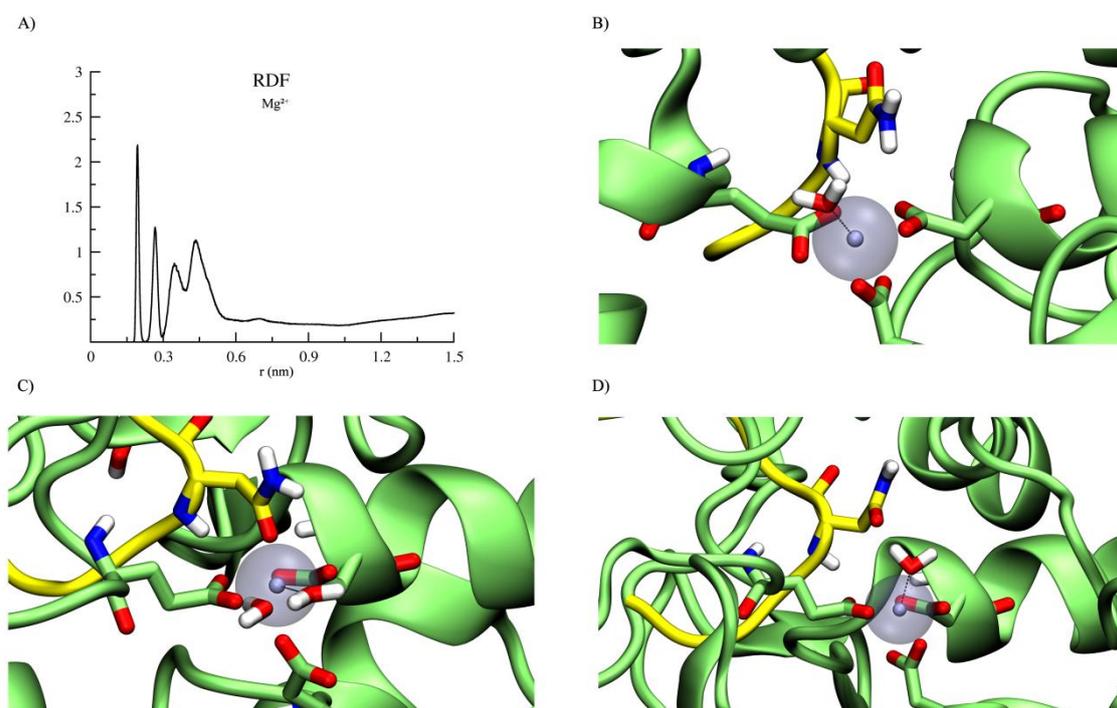


Figure 6. A) RDF analysis demonstrating the presence of solvent molecules near the catalytic ion (less than 0.3 nm). B-D) Snapshots of the simulation demonstrating all geometries of different interactions between the residues from the catalytic site with the divalent metal and the water molecules.

Regarding the glycan substrate, we modeled the oligosaccharide chain from *C. jejuni*, since this the most studied system in Bacteria<sup>16</sup> and previous NMR studies have

characterized its structure (PDB ID: 2K33),<sup>17</sup> allowing us to compare with another experimental data. Still, PglB has a promiscuous behavior, as the only pre-requisite for the glycan substrate to be transferred is the presence of the N-acetylation at the position C2.<sup>18</sup> Thus, by employing the oligosaccharide chain structure contained in the NMR structure from 2K33, we were able to analyze and compare the glycosidic linkages  $\phi$  and  $\psi$  dihedral angles with the available data, further expanding this comparison to the conformations of the linkage inside the proteins cavities and in a full protein in solution (Table 2). As demonstrated, the average conformational profile of the oligosaccharide chain attached to the peptide in the catalytic site remains mostly rigid during the simulation, populating similar geometries, with minor fluctuations to other conformational states. The same behavior is observed for the NMR angles evaluated, with a few exceptions. Two angles  $\psi$  demonstrated slight variations ( $-133\pm 1$  to  $-93\pm 8$  and  $-134\pm 5$  to  $-95\pm 9$ ). One major difference is observed for the  $\beta$ -D-Bac-(1 $\rightarrow$ N)-Asn linkage  $\phi$  angle, showing an average of  $111\pm 9$  at the simulation and  $-51\pm 98$  at the NMR data. This can be explained by the huge variations within the NMR detected geometries. By dividing in three smaller subsets of values, we can obtain -146, -115 and 63 values for the  $\phi$  angle, exhibiting multiple conformations, in which one of them agrees with the data found in our simulation. Therefore, we assume that *C. jejuni* possess a stable rod shape N-glycan chain, due to its structure occupation of its minimal energy regions that allows for this pattern.

**Table 2.** Dihedral angles of the glycosidic linkages from the N-glycan of *C. jejuni*, obtained from the 200ns MD simulation data.

Glycosidic Linkage	Dihedral angle (°)			
	200ns simulation		NMR data	
	$\phi$	$\psi$	$\phi$	$\psi$
$\beta$ -D-Bac-(1 $\rightarrow$ N)-Asn	111 $\pm$ 9	-152 $\pm$ 34	-51 $\pm$ 98	177 $\pm$ 22
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 3)- $\beta$ -D-Bac	74 $\pm$ 7	-137 $\pm$ 7	83 $\pm$ 5	-134 $\pm$ 5
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 4)- $\alpha$ -L-GalNAc	74 $\pm$ 9	-127 $\pm$ 12	87 $\pm$ 1	-140 $\pm$ 2
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 4)- $\alpha$ -L-GalNAc	97 $\pm$ 8	-93 $\pm$ 8	87 $\pm$ 8	-133 $\pm$ 1
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 4)- $\alpha$ -L-GalNAc	68 $\pm$ 10	-148 $\pm$ 17	91 $\pm$ 5	-138 $\pm$ 6
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 4)- $\alpha$ -L-GalNAc	92 $\pm$ 9	-95 $\pm$ 9	82 $\pm$ 6	-134 $\pm$ 5
$\alpha$ -L-GalNAc-(1 $\rightarrow$ 3)- $\beta$ -D-Glc	-66 $\pm$ 18	-129 $\pm$ 26	-76 $\pm$ 4	-159 $\pm$ 7

## Methods

### Nomenclature and software.

The IUPAC<sup>19</sup> proposed recommendations for nomenclature and symbols were used. Regarding MD simulations, the GROMACS 4.5.1 simulation suite<sup>20</sup> was employed, along with the GROMOS 53A6 force field<sup>21</sup> for the systems without the glycan chain and the GROMOS 53A6<sub>GLYC</sub> force field<sup>22</sup> for the system presenting the glycan chain. The adjacent carbohydrate relative orientations are described by the dihedral angles  $\phi$  e  $\psi$  which pertains to their glycosidic linkages. Each linkage corresponding to (1 $\rightarrow$ X), where X is ‘3’ or ‘4’, has its  $\phi$  e  $\psi$  angles defined as shown below:

$$\phi = O5 - C1 - OX - CX \quad (1)$$

$$\psi = C1 - OX - CX - C(X-1) \quad (2)$$

For the manipulation and visualization of structures, the softwares VMD,<sup>23</sup> and PyMOL<sup>24</sup> were employed.

### **System preparation.**

The crystallographic structure of the *C. lari* PglB was retrieved from the Protein Data Bank (PDB), under the code 3RCE.<sup>8</sup> There were two regions with missing atoms in the structure: one located at the TM domain, comprehending 25 residues from 283 to 306 and residing at the external loop 5 (EL5), while the other is located at the PP domain, comprehending 3 residues from 605 to 607 and residing at the  $\beta$ -barrel like structure. Using the web server FALC-Loop,<sup>25</sup> we were able to fill both loops. Since the EL5 is considered an important structure, due to possible connections to the catalysis mechanism, extra care was taken when choosing an appropriate model for the loop modeling. Based on stereochemistry analysis by submitting the structures to PDBSUM server, which uses the PROCHECK software,<sup>26</sup> we chose the model that the residues position did not block the hydrophobic groove where the lipid-linked oligosaccharide (LLO) chain is supposed to interact. Departing from this starting structure, three systems were studied: PglB at the unbound state; PglB with a bound acceptor peptide (already co-crystallized in the 3RCE structure); PglB with a bound glycosylated peptide with the heptasaccharide chain from *C. jejuni*. The structure for the oligosaccharide chain was extracted from the PDB, under the code 2K33.

### **Membrane insertion of PglB.**

Regarding the insertion of the protein in a membrane, the employed protocol followed the InflateGRO methodology.<sup>27</sup> Briefly, this consists in inserting PglB into a pre-equilibrated bilayer patch, with lipids overlapping the protein, and then expanding the dimensions of the box, as well as, translating all lipids laterally, so that no more overlap is found. After that, a series of minimizations steps occur, along with the box dimensions being compressed and the lipids translated back to the center of the system, until the system reaches the desired density. During all this process, the protein is under strong position restraints, so that the structure is not affected. A bilayer model constituted of 490 POPE lipids was obtained at the end of the protocol, as some lipids had to be removed from the initial patch of 512 lipids because of overlap issues. Parameters accounting for the POPE membrane parameterization are presented in the *Supplementary Material*.

### **Molecular Dynamics Simulations.**

Following the membrane insertion steps, the rectangular box was then solvated with SPC water model<sup>28</sup> and periodic boundary conditions. Before this process, we did a common procedure where the van der Waals radius of the C atom was raised from 0.15 to 0.375, aiming to avoid the water molecules filling any left space between the lipids and the protein. After solvation, this parameter was returned back to its original value. Counterions of Na<sup>+</sup> were added to neutralize the systems, when needed. The LINCS algorithm<sup>29</sup> was chosen to constrain covalent bond lengths. This way, an integration step of 2 fs was applied. As for the electrostatic interactions, calculations were performed by the Particle Mesh Ewald (PME) method.<sup>30</sup> The pressure barostat chosen was Parrhinello-Rahman,<sup>31,32</sup> with a 2.0 ps coupling constant, while the temperature thermostats chosen was V-rescale<sup>33</sup> (NVT step) and Nosé-Hoover (NPT equilibration and production MD),<sup>34,35</sup> with a coupling constant of  $\tau=0.5$ . Additionally, semiisotropic

pressure coupling type was applied, accounting for the presence of the membrane. A constant temperature of 310 K was also implemented. Steepest Descent algorithm was used in the energy minimizations performed. First, two simulations of equilibration were performed with position restraints: a NVT and a NPT of 0.1 ns and 0.9 ns, respectively. Subsequently, 200 ns NPT MD simulations were performed for each of the four systems, generating the production run where data was collected for the systems analysis.

### References.

1. Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 3, 97-130.
2. Abu-Qarn, M., Yurist-Doutsch, S., Giordano, A., Trauner, A., Morris, H. R., Hitchen, P., Medalia, O., Dell, A. & Eichler, J. (2007). *Haloferax volcanii* AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J. Mol. Biol.* 14, 1224–1236.
3. Eichler, J. (2013). Extreme sweetness: protein glycosylation in archaea. *Nat. Rev. Microbiol.* 11, 151-156.
4. Szymanski, C. M., Burr, D. H. & Guerry, P. (2002). Campylobacter protein glycosylation affects host cell interactions. *Infect. Immun.* 70, 2242–2244.
5. Nothaft, H. & Szymanski, C. M. (2013). Bacterial protein N-glycosylation: New perspectives and applications. *J. Biol. Chem.* 2013, 288, 6912-6920.
6. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* 32, 1022–1030.

7. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W. & Aebi, M. (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298, 1790–1793.
8. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. *Nature*. 474, 350–355.
9. Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita, T., Maenaka, K. & Kohda, D. (2008). Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* 27, 234–243.
10. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J. & Kohda, D. (2010). Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.* 285, 4941–4950.
11. Jaffee, M. B. & Imperiali, B. (2011). Exploiting Topological Constraints To Reveal Buried Sequence Motifs in the Membrane-Bound N-Linked Oligosaccharyl Transferases. *Biochemistry*. 50, 7557–7567.
12. Matsumoto, S., Igura, M., Nyirenda, J., Matsumoto, M., Yuzawa, S., Noda, N., Inagaki, F. & Kohda, D. (2012). Crystal Structure of the C-Terminal Globular Domain of Oligosaccharyltransferase from *Archaeoglobus fulgidus* at 1.75 Å Resolution. *Biochemistry*. 51, 4157-4166.
13. Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., Catrein, I. & Aebi, M. (2006). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science*. 314, 1148–1150.
14. Ihssen, J., Kowarik, M., Wiesli, L., Reiss, R., Wacker, M. & Thöny-Meyer, L. (2012). Structural insights from random mutagenesis of *Campylobacter jejuni* oligosaccharyltransferase PglB. *BMC Biotechnol.* 12, 67-79.

15. Katz, A. K., Glusker, J. P., Beebe, S. A. & Bock, C. W. (1996). Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *J. Am. Chem. Soc.* 118, 5752-5763.
16. Nothaft, H. & Szymanski, C. M. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* 8, 765-778.
17. Slynko, V., Schubert, M., Numao, S., Kowarik, M., Aebi, M. & Allain, F. H. (2009). NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J. Am. Chem. Soc.* 131, 1274-1281.
18. Wacker, M., Feldman, M. F., Callewaert, N., Kowarik, M., Clarke, B. R., Pohl, N. L., Hernandez, M., Vines, E. D., Valvano, M. A., Whitfield, C. & Aebi, M. (2006) Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci.* 103, 7088–7093.
19. IUPAC-IUBMB. (1996). Nomenclature of Carbohydrates. Recommendations. *Pure Appl. Chem.* 68, 1919-2008.
20. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. & Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 29, 845-854.
21. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25, 1656-1676.

22. Pol-Fachin, L., Rusu, V. H., Verli, H. & Lins, R. D. (2012) GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theo. Comp.* 8, 4681-4690.
23. Humphrey, W., Dalke, A. & Schulten, K. (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graphics.* 14, 33-38.
24. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, Schrodinger Sales Center, Portland, OR, USA.
25. Ko, J., Lee, D., Park, H., Coutsias, E. A., Lee, J. & Seok, C. (2011) The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res.* 39, W210-W214.
26. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.
27. Kandt, C., Ash W. L., Tieleman D. P. (2007) Setting up and running molecular dynamics simulations of membrane proteins. *Methods.* 41, 475-488.
28. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. J. (1987) The Missing Term in Effective Pair Potentials. *Phys. Chem.* 91, 6269-6271.
29. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463-1472.
30. Darden, T., York, D. & Pedersen, L. (1993) Particle mesh Ewald: An  $N \bullet \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089-10092.
31. Parrinello, M. & Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182-7190.
32. Nosé, S. & Klein, M. L. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50, 1055-1076.

33. Bussi, G., Donadio, D. & Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101-014107.
34. Nosé, S. (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52, 255-268.
35. Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31, 1695-1697.

## 4. 2 FILOGENIA

Diversas abordagens foram realizadas buscando a obtenção de dados estatisticamente confiáveis para a inferência filogenética, tanto realizando análises para os domínios separadamente como contendo os três juntos. Utilizando MV e EM no alinhamento contendo sequências dos três domínios, foram obtidas três árvores diferentes (Figura 9 e 10). O padrão observado nessas análises aponta para o domínio Bacteria (violeta) como o mais ancestral, demonstrando proximidade a algumas metanógenas (ciano) do domínio Archaea, como as pertencentes aos gêneros *Methanothermobacter*, *Methanobrevibacter*, *Methanocaldococcus* e *Methanococcus*. Posteriormente, mais dois grupos são identificados dentro o domínio Archaea, um grande grupo compreendendo mais membros do filo Euryarchaeota (verde), além das metanógenas citadas, e um grupo com menos representantes contendo os membros do filo Crenarchaeota (amarelo), sendo o primeiro mais ancestral e o segundo mais próximo do domínio Eukarya. (vermelho) Contudo, essa relação não pode ser inferida categoricamente, pois o suporte estatístico obtido não possui valores adequados. As sequências selecionadas do domínio Eukarya (vermelho) são representadas na árvore, primeiramente, pelos Protistas, pelas Algae e por alguns membros de Plantae mais primitivas. A seguir, surgem os representantes de Animalia e depois o Fungi, esses os organismos com as OSTs mais recentes nessas árvores. Ainda, as sequências mais próximas de Fungi são aquelas denominadas como do subtipo STT3B, uma isoforma da STT3A de eucariotos.

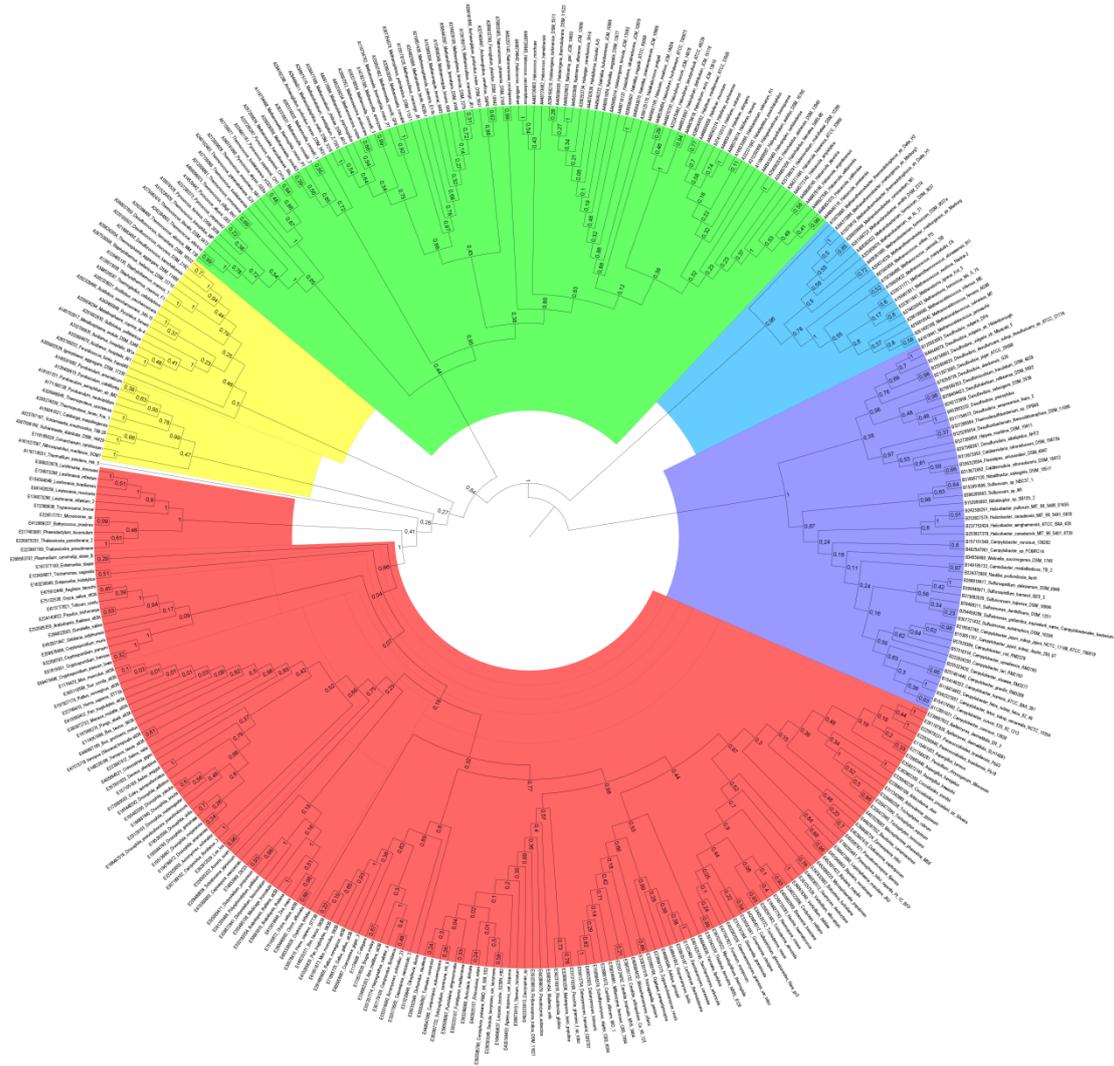


Figura 9. Árvore filogenética envolvendo os três domínios da vida e obtida pelo método de EM. Destacadas estão as bactérias (violeta), *archaeas* metanógenas (ciano), euryarcheota (verde), crenarcheota (amarelo) e eucariotos (vermelho).

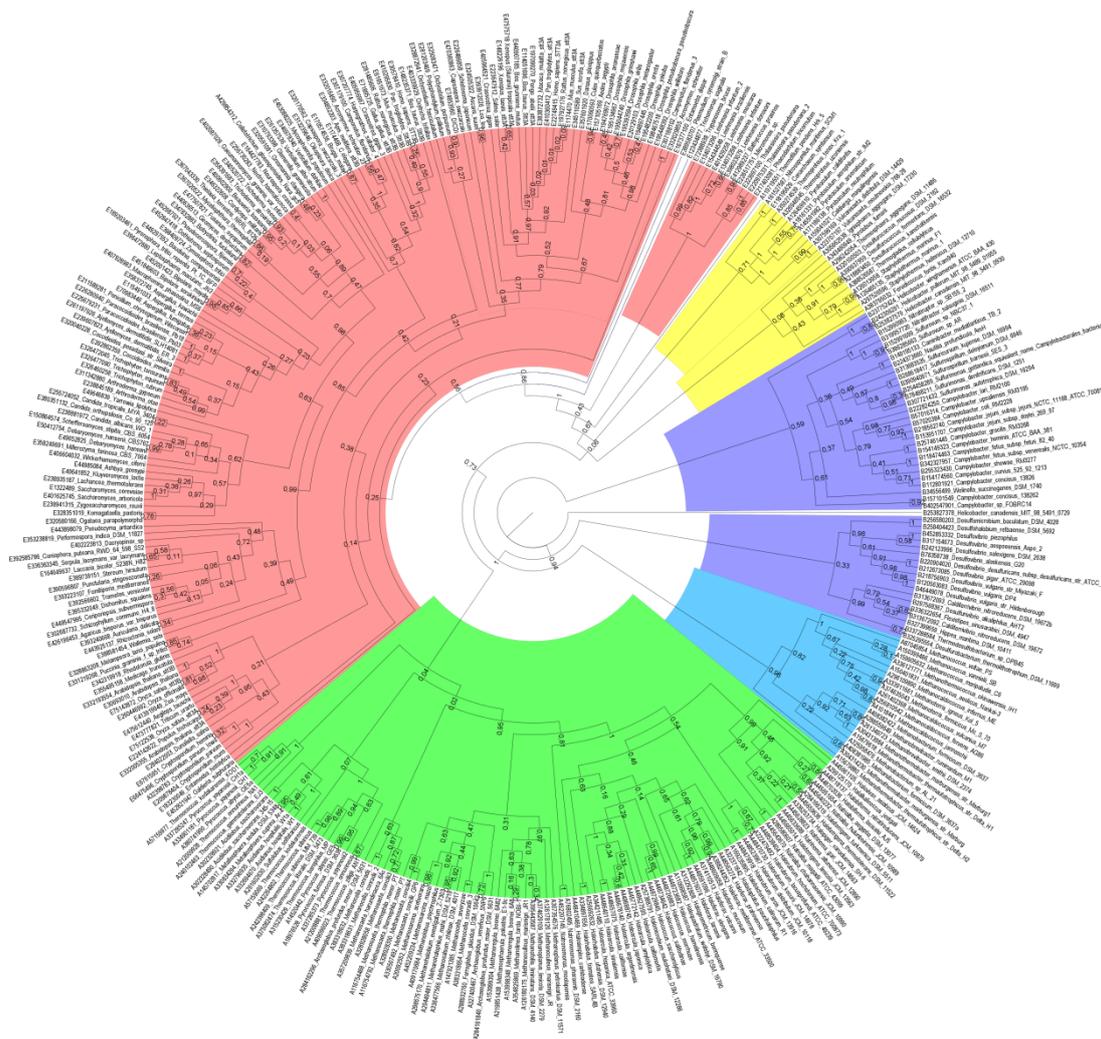


Figura 10. Árvore filogenética envolvendo os três domínios da vida e obtida pelo método de ML. Destacadas estão as bactérias (violeta), archaeas metanógenas (ciano), euryarcheota (verde), cenarcheota (amarelo) e eucariotos (vermelho).

O domínio Bacteria foi analisado, objetivando elucidar cada domínio separadamente, buscando assim atingir árvores com suporte estatístico apropriado e avaliar pormenorizadamente os domínios. Para tanto, o método de Inferência Bayesiana (IB) foi aplicado para as sequências de bactérias, além da inclusão de três sequências eucarióticas como grupo externo (Figura 11). O suporte estatístico mostrou-se apropriado e dessa forma foi possível observar a presença do motivo MIV de *C. jejuni* e a sua transição, a partir dos organismos mais ancestrais, começando com IKW e seguido por MKI, MLI, ILI e MII. Os organismos que apresentam MII pertencem aos gêneros *Helicobacter* e *Campylobacter*, membros da mesma ordem (Campylobacterales). Por fim, o

MIV é encontrado tanto no gênero *Campylobacter*, como em *Sulfurimonas*, *Sulfurovum*, *Sulfuricurvum*, *Sulfuruspirillum*, *Wolinella*, *Caminibacter*, *Nautilia*, *Nitratiruptor* e *Nitratifactor*, sendo a variante mais difundida entre as sequências disponíveis até o presente momento. A enzima de *C. lari*, estudada nesse trabalho, possui a variante MIV e está disposta no mesmo grupo, porém anterior a *C. jejuni*. Segundo a descrição encontrada nessa árvore, os grupos mais ancestrais que possuem a N-Glicosilação são os compreendendo membros do filo Proteobacterias da classe Deltaproteobacterias e membros de outros filios, como o Thermodesulfobacteria, o Deferribacteres e o Aquificae. A classe Epsilonproteobacteria, que possui o maior número de representantes já descritos, é caracterizada como sendo posterior aos supracitados.

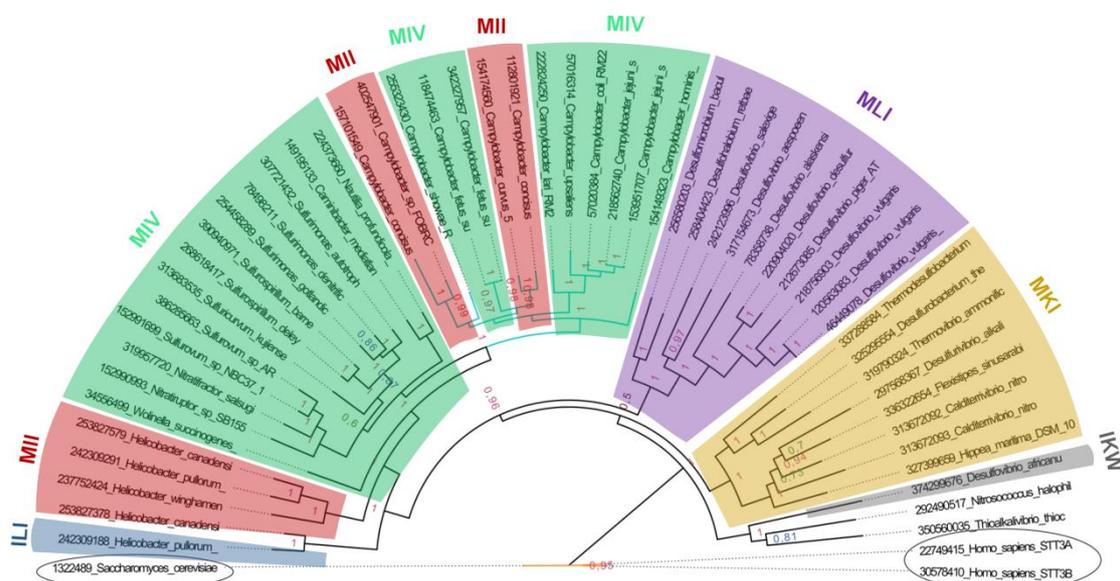


Figura 11. Árvore filogenética obtida pelo método de IB e compreendendo espécies do domínio Bacteria e três sequências de grupo externo. Os grupos estão coloridos de acordo com a variante do motivo MIV apresentado. Em dois casos não foi possível identificar qual variação os organismos apresentavam.

## 5. DISCUSSÃO GERAL

O entendimento do processo de N-glicosilação, assim como a compreensão de sua via e seus componentes proporciona novos caminhos tanto nas buscas envolvendo aplicações na terapêutica como na biotecnologia. Muitos avanços ocorreram nos últimos anos, entretanto, diversas questões ainda permanecem em aberto. No contexto das potenciais aplicações e no uso em larga escala da PglB, ainda é difícil prever quais sítios serão glicosilados e os possíveis efeitos que essas glicosilações causariam na proteína (Baker *et al.*, 2013). Ainda, os protocolos da sua expressão e purificação pecam em eficiência, especialmente devido a sua inserção em membrana e instabilidade, apesar de avanços recentes nesse âmbito (Jaffee & Imperiali, 2013).

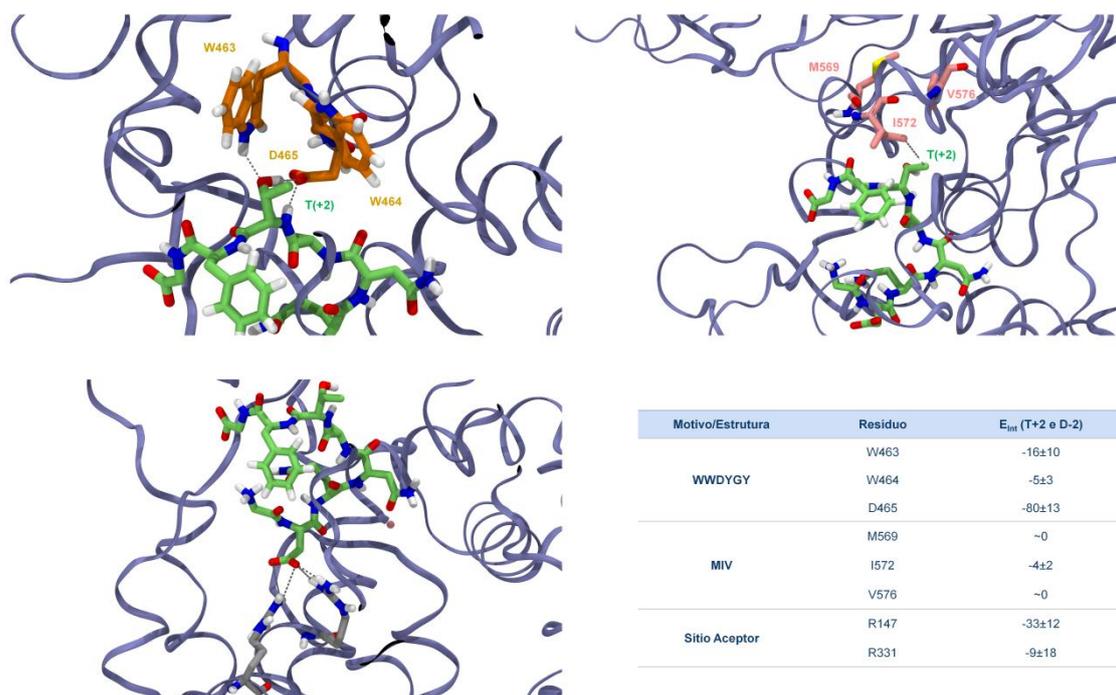
No que diz respeito a estudos estruturais, a despeito de já possuir sua estrutura elucidada, não se tem certeza da localização do sítio onde o lipídeo carreador e seu oligossacarídeo irão aportar. A estrutura cristalográfica também não fornece informações suficientes envolvendo a alça EL5, já que uma porção de 25 resíduos está incompleta devido ao elevado grau de desordem pertencente a essa região. Como discutido previamente, foi predito o envolvimento dessa alça na catálise, formando o sítio catalítico a partir da interação com a proteína aceptora (Lizak *et al.*, 2011). A modelagem dessa estrutura nos modelos construídos no presente trabalho permitiu uma primeira análise em relação a ela. Foi possível observar nas simulações que, de fato, a EL5 possui maior flexibilidade na sua forma nativa, particularmente em suas porções inicial e final, em relação às outras duas formas que possuem os substratos complexados. Contudo, não foi observado um movimento tão amplo, a ponto de desorganizar o sítio catalítico. Nesse caso deve ser levado em consideração o fato de que a estrutura cristalina foi obtida na sua forma ativa, ou seja, complexada ao substrato proteico. Esses contatos cristalográficos podem interferir no resultado da simulação, visto que determinadas interações já estarão transcorrendo. Em princípio, a simulação da PglB em sua forma nativa mostraria essa movimentação, porém essa pode ser uma característica que depende de um longo tempo para acontecer. O tempo de 200ns para as simulações não necessariamente é suficiente para se observar todos as

particularidades desse sistema, pois diferentes casos possuem diferentes complexidades e exigirão magnitudes de tempo próprias.

A dificuldade em se obter a estrutura de uma proteína transmembrana é um dos principais fatores impeditivos da expansão do conhecimento acerca das diversas OSTs. Diversas estruturas de OSTs já foram cristalizadas, entretanto todas elas envolvem apenas o domínio solúvel (periplasmático) da proteína. Essas estruturas serviram de base para o avanço do conhecimento estrutural e definição de motivos importantes para a N-glicosilação, como o WWDYGY, o MIV e seus variantes, além do recente DxxK. Esses motivos possuem um papel relevante, pois são necessários no correto reconhecimento da assinatura da N-glicosilação e possuem implicações na catálise. Nas simulações realizadas foi observado que parte do motivo WWD manteve-se firmemente interagindo com o aminoácido da posição +2 da assinatura (T) (Figura 12), concordando com os dados da literatura (Lizak *et al.*, 2011). Nenhuma inferência foi realizada para a porção YG, enquanto a Tyr468, previamente descrita como interagindo com o oligossacarídeo do sítio (Lizak *et al.*, 2011), não foi detectada como importante nessa relação. Quanto ao motivo MIV, que possui a Ile572 como relevante na preferência entre serina ou treonina na posição +2 da assinatura, não foi possível identificar interações fortes entre esse resíduo e a metila da cadeia lateral do resíduo de treonina (Figura 12). Por fim, o motivo DxxK apresentou interação direta com elementos do sítio catalítico, como a asparagina aceptora e os resíduos Bac e GalNAc1. Esses dados indicam uma maior contribuição desse motivo no processo de N-glicosilação, um fato que pode ser estudado mais amplamente a partir de ensaios de mutagênese. Dados prévios (Jaffee & Imperiali, 2011; Ihssen *et al.*, 2012) indicam que mutações em alguns pontos desse motivo causam diminuição na atividade da enzima, todavia sem causar perda de função. Adicionalmente, os dados obtidos a partir das simulações sugerem o envolvimento de mais elementos estruturais no auxílio dos mecanismos relacionados à catálise. Por exemplo, a alça D<sub>481</sub>xxK<sub>484</sub> manteve-se próxima a Asn aceptora durante a simulação, inclusive interagindo a partir de duas Gly (482 e 483) e uma His (485) com o átomo de nitrogênio da amida durante alguns momentos. Ainda, outros resíduos do domínio transmembrana mantiveram-se próximos ao sítio e interagindo com alguns aminoácidos ali

presentes, tais como a Thr54, a Thr148, a Tyr152 e a Gln301. Dessa maneira, estudos por mutagênese poderiam contribuir para a compreensão da relevância tanto estrutural como funcional do papel desses aminoácidos na PglB.

No domínio transmembrana estão localizados os resíduos do sítio catalítico, além de pelo menos mais três resíduos de arginina já descritos como importantes no reconhecimento tanto do substrato proteico (Arg147 e Arg331) como do substrato sacarídico (Arg375), e mais um (Arg31) que foi identificado por mutações como essencial para a atividade da PglB, apesar de distante do sítio catalítico. A Arg331 é conservada somente no domínio Bacteria e, pela proximidade no cristal com o aminoácido da posição -2 da assinatura (D/E), implicado como identificador dessa posição (Figura 9). Nas simulações, ele manteve esse papel, porém com menor intensidade que a Arg147, sendo esse considerado relevante na coordenação de resíduos presentes no sítio catalítico. É importante salientar que na estrutura cristalina, apesar de próximo dos resíduos Glu319 e Asp156, a Arg147 está próxima também do Asp(-2), indicando que podem ocorrer interações nessa região. Esse aminoácido, então, pode realizar uma função dupla tanto na coordenação de resíduos catalíticos como auxiliando na manutenção do substrato proteico firmemente interagindo. Dessa maneira, a importância da obtenção da estrutura transmembrana para os estudos de estrutura e função das OSTs é um fator que não pode ser ignorado.



**Figura 12.** Descrição das interações de motivos conservados com aminoácidos importantes para o processo de N-glicosilação, identificadas a partir das simulações de DM.

A estrutura cristalina disponível propõe uma rede de coordenação de ligações de hidrogênio que explica como um nucleófilo fraco como as amidas aumentariam sua eletronegatividade e realizariam o ataque nucleofílico no carbono da posição 1 da bacilosamina. A partir das simulações realizadas, foi possível destacar diferentes estados conformacionais compatíveis com a coordenação do metal  $Mg^{2+}$  pelos resíduos catalíticos. Um desses estados observados corresponde ao encontrado no cristal, mas como um estado transiente durante o cálculo. Deste modo, podemos propor que as posições ótimas para a catálise do processo de transferência da glicana N-ligada acontecem em subestados da simulação, enquanto outras coordenações ocorrem mais prevalentemente no resto do tempo. Foram observadas tanto distâncias compatíveis com coordenações relativas a moléculas de água como envolvendo a carbonila da cadeia lateral da asparagina aceptora, e essas mantinham a estrutura octaédrica (preferencial dos íons de  $Mg^{2+}$ ) apesar de, em alguns momentos, mais de 6 átomos estarem próximos o suficiente para coordenar o metal.

Finalmente, algumas ressalvas são necessárias quanto ao dado cristalográfico. Apesar de sua enorme importância na descoberta dos mecanismos de reconhecimento do substrato e na identificação dos resíduos vitais para o processo de catálise da transferência da glicana para proteína acceptora, a resolução dele é de 3,4 Å. Esse valor é considerado alto e pode propiciar erros na interpretação da posição de certas cadeias laterais, por exemplo. Esse fato poderia ser uma justificativa para a diferença encontrada na simulação entre a posição do grupo amida e da carbonila da Asn-acceptora. Outros fatores relevantes envolvem o alto valor de pH (9,4) em que ocorreu a cristalização, pois essa é uma faixa onde a atividade da enzima é muito baixa. (Jaffee & Imperiali, 2011) Portanto, estudos adicionais ainda são necessários para o desenvolvimento de hipóteses mais consistentes quanto à elucidação desses mecanismos sem resposta.

A correta compreensão desses aspectos das OSTs, combinada a estudos envolvendo as enzimas da via de N-glicosilação em bactérias como *C. jejuni*, somado a consequente transfecção dessa maquinaria para organismos modelo como *E. coli*, apresentam-se como alternativas valiosas para a produção em larga escala de glicoproteínas. Nessa linha, estudos filogenéticos podem não só contribuir para a compreensão da ancestralidade dessa via nos domínios da vida, mas também para a realização de análises comparativas entre a conservação de motivos importantes para a catálise e para o reconhecimento dos substratos. A conjugação dos dados estruturais e filogenéticos permite que seja avaliada a possibilidade de extrapolação dos dados para outros organismos, fato que pode contribuir para o entendimento das OSTs em outros domínios. As estruturas das OSTs podem variar bastante em termos de tamanho de sequência, sendo que as eucarióticas ainda possuem diversas subunidades adicionais. Ainda, diversas unidades estruturais já foram identificadas como presentes nos diversos organismos.

Os dados preliminares de filogenia da PglB e suas equivalentes nos outros domínios convergem para esse mesmo caminho de variabilidade, tornando difícil a resolução da história evolutiva dessas enzimas. Aparentemente, os motivos conservados são exceções dentro dessas sequências, que demonstram baixíssima identidade entre si, e os métodos aplicados não foram suficientes para proporcionar confiabilidade grande aos

dados obtidos. Outras abordagens foram testadas como, por exemplo, a utilização de alinhamentos contendo apenas a região periplasmática, os motivos conservados, a exclusão das regiões hipervariáveis e a utilização somente da porção contendo a unidade estrutural CC, já que essa está presente em todas as estruturas já elucidadas. Entretanto, nenhum desses métodos conseguiu propiciar um aumento significativo nos valores de confiabilidade, especialmente nas sequências relativas aos domínios Eukarya e Archaea. Apesar disso, esse trabalho demonstra a primeira análise filogenética envolvendo os três domínios da vida em que o suporte estatístico é mostrado e utilizando um método como a MV. Previamente, foi proposto que o domínio Bacteria teria recebido a N-Glicosilação via transferência horizontal de genes de Archaea, advindo de metanógenas (Maita *et al.*, 2010). A árvore encontrada demonstra certa proximidade entre esses ramos, o que pode ser um indicativo importante da ocorrência desse processo. Em geral, essas metanógenas apresentam variantes do motivo MIV, como o MKW e o MIW, um dado que reforça essa informação. De qualquer forma, novas abordagens devem ser testadas para que seja possível inferir adequadamente a ancestralidade dessa proteína.

## 6. CONCLUSÕES

A partir dos objetivos previamente descritos, a concretização desse trabalho tornou possível:

- Realizar uma análise do sítio catalítico da enzima PglB, descrevendo o comportamento dos resíduos do sítio ativo, bem como avaliar a movimentação desses aminoácidos em relação ao íon  $Mg^{2+}$  e a presença de moléculas de água, mostrando as diversas configurações encontradas durante a simulação.
- Estudar as interações dos motivos conservados, importante para o reconhecimento dos substratos e do processo de catálise, discriminando qualitativamente a intensidade dessas interações e fornecendo mais *insights* sobre esses resíduos.
- Identificar movimentos estruturais importantes que apontam para uma abertura da cavidade da proteína aceptora, ao passo que as alças circundando a cavidade onde é posicionada a cadeia oligossacarídica realizam um movimento de fechamento, pressionando o carboidrato.
- Inferir as relações de ancestralidade compreendendo o domínio Bacteria, com suporte estatístico adequado, a partir do método de Inferência Bayesiana e observar a transição do motivo MIV e suas variantes.
- Realizar uma primeira caracterização filogenética envolvendo os três domínios com um número expressivo de sequências, onde, apesar do baixo suporte estatístico, é possível observarmos relações constantes mesmo utilizando métodos diferentes (MV e EM).

Os resultados desse trabalho apontam para a caracterização estrutural e dinâmica da PglB e contribuem para o entendimento, em nível molecular, dos mecanismos envolvendo as OSTs bacterianas. Esses dados potencialmente serão extrapolados para outras OSTs, visto que essas enzimas, apesar de não possuírem grande identidade de sequência, demonstram topologias transmembranas similares, assim como formas de reconhecimento de substratos e o processo da catálise.

## 7. PERSPECTIVAS

A partir do trabalho realizado, novas abordagens podem ser aplicadas buscando expandir o delineamento inicial do trabalho:

- Realizar cálculos de mecânica quântica, a partir de diferentes conformações, a fim de avaliar mais precisamente a coordenação ocorrendo no sítio catalítico.
- Parametrizar a undecaprenila difosfatada, visando a influência que a ligação desse lipídio carreador no carboidrato irá provocar, bem como utilizá-lo complexado com a PglB, observando assim suas interações e possíveis alterações na estrutura da enzima.
- Simular a glicosilação da PglB, já que não há estudo algum relativo a essa informação.
- Realizar modelagem comparativa, utilizando os cristais existentes das regiões periplasmáticas de OSTs para a obtenção de modelos completos das enzimas de Archaea e Eukarya.
- Buscar outras maneiras de realizar a inferência filogenética da OST com os três domínios da vida, almejando a obtenção de dados estatisticamente confiáveis.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

- Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J.: Basic local alignment search tool. *J. Mol. Biol.*, **1990**, 215, 403-410.
- Apweiler, R.; Hermjakob, H.; Sharon, N.: On the frequency of protein glycosylation, as deduced from the analysis of the SWISS-PROT database. *Biochim. Biophys. Acta*, **1999**, 1473, 4-8.
- Baker, J. L.; Çelik, E.; DeLisa, M. P.: Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. *Trends Biotechnol.*, **2013**, 31, 313-323.
- Bussi, G.; Donadio, D.; Parrinello, M.: Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **2007**, 126, 014101-014107.
- Calo, D.; Kaminski, L.; Eichler, J.: Protein glycosylation in archaea: Sweet and extreme. *Glycobiology*, **2010**, 20, 1065–1076.
- Chen, M. M.; Weerapana, E.; Ciepichal, E.; Stupak, J.; Reid, C. W.; Swiezewska, E.; Imperiali, B.: Polyisoprenol Specificity in the *Campylobacter* Jejuni N-linked Glycosylation Pathway. *Biochemistry*, **2007**, 46, 14342–14348.
- Darden, T.; York, D.; Pedersen, L.: Particle mesh Ewald: An  $N \bullet \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.*, **1993**, 98, 10089–10092.
- DeLano, W. L.: The PyMOL Molecular Graphics System, Schrodinger Sales Center, Portland, OR, USA, **2002**. <http://www.pymol.org>.
- Dempski Robert, E., Jr.; Imperiali, B.: Oligosaccharyl transferase: Gatekeeper to the secretory pathway. *Curr. Opin. Chem. Biol.*, **2002**, 6, 844–850.
- Feldman, M.F.; Wacker, M.; Hernandez, M.; Hitchen, P. G.; Marolda, C. L.; Kowarik, M.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M.: Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci.*, **2005**, 102, 3016–3021.
- Focareta, A. Paton, J. C.; Morona, R.; Cook, J.; Paton, A. W.: A recombinant probiotic for treatment and prevention of cholera. *Gastroenterology*, **2006**, 130, 1688–1695.

- Glover, K. J.; Weerapana, E.; Imperiali, B.: In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, *102*, 14255–14259.
- Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B.: Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry*, **2006**, *45*, 5343–5350.
- Grass, S.; Lichti, C. F.; Townsend, R. R.; Gross, J.; St Geme, J. W. 3<sup>rd</sup>: The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog.*, **2010**, *6*, e1000919
- Helenius, J.; Ng, D. T. W.; Marolda, C.; Walter, P.; Valvano, M. A.; Aebi, M.: Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature*, **2002**, *415*, 447–450.
- Herscovics, A.: Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochim. Biophys. Acta*, **1999**, *1473*, 96–107.
- Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M.: LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **1997**, *18*, 1463-1472.
- Hoover, W. G.: Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.*, **1985**, *31*, 1695-1697.
- Huelsenbeck, J. P.; Ronquist, F.: MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **2001**, *17*, 754–755.
- Hug, I.; Zheng, B.; Reiz, B.; Whittal, R. M.; Fentabil, M. A.; Klassen, J. S.; Feldman, M. F.: Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. *J. Biol.Chem.*, **2011**, *286*, 37887–37894.
- Humphrey, W.; Dalke, A.; Schulten, K.: VMD - Visual Molecular Dynamics. *J. Molec. Graphics.*, **1996**, *14*, 33-38.
- Ielmini, M. V.; Feldman, M. F.: *Desulfovibrio desulfuricans* PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology*, **2011**, *21*, 734–742.

- Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D.: Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.*, **2008**, *27*, 234–243.
- Ihssen, J.; Kowarik, M.; Dilettoso, S.; Tanner, C.; Wacker, M.; Thöny-Meyer, L.: Production of glycoprotein vaccines in *Escherichia coli*. *Microb. Cell Fact.*, **2010**, *9*, 61-73.
- Ihssen, J.; Kowarik, M.; Wiesli, L.; Reiss, R.; Wacker, M.; Thöny-Meyer, L.: Structural insights from random mutagenesis of *Campylobacter jejuni* oligosaccharyltransferase PglB. *BMC Biotechnol.*, **2012**, *12*, 67-79.
- Ilg, K.; Yavuz, E.; Maffioli, C.; Priem, B.; Aebi, M.: Glycomimicry: display of the GM3 sugar epitope on *Escherichia coli* and *Salmonella enterica* sv *Typhimurium*. *Glycobiology*, **2010**, *20*, 1289–1297.
- Iwashkiw, J.A.; Fentabil, M. A.; Faridmoayer, A.; Mills, D. C.; Peppler, M.; Czibener, C.; Ciocchini, A. E.; Comerci, D. J.; Ugalde, J. E.; Feldman, M. F.: Exploiting the *Campylobacter jejuni* protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis. *Microb. Cell Fact.*, **2012**, *11*, 13-23.
- Jaeken, J.: Congenital disorders of glycosylation (CDG): it's (nearly) all in it! *J. Inherit. Metab. Dis.*, **2011**, *34*, 853–858.
- Jaffee, M. B.; Imperiali, B.: Exploiting Topological Constraints To Reveal Buried Sequence Motifs in the Membrane-Bound N-Linked Oligosaccharyl Transferases. *Biochemistry*, **2011**, *50*, 7557–7567.
- Jaffee, M. B.; Imperiali, B.: Optimized protocol for expression and purification of membrane-bound PglB, a bacterial oligosaccharyl transferase. *Protein Expr Purif.*, **2013**, *89*, 241-50.
- Jelk, J.; Gao, N.; Serricchio, M.; Signorell, A.; Schmidt, R. S.; Bangs, J. D.; Acosta-Serrano, A.; Lehrman, M. A.; Bütikofer, P.; Menon, A. K.: Glycoprotein biosynthesis in a eukaryote lacking the membrane protein Rft1. *J. Biol. Chem.*, **2013**, *288*, 20616-20623.
- Johansen, P.G.; Marshall, R.D.; Neuberger, A.: Carbohydrates in protein. The preparation and some of the properties of a glycopeptide from hen egg ovalbumin. *Biochem. J.*, **1961**, *78*, 518–527.
- Kaminski, L.; Guan, Z.; Abu-Qarn, M.; Konrad, Z.; Eichler, J.: AglR is required for addition of the final mannose residue of the N-linked glycan

- decorating the *Haloferax volcanii* S-layer glycoprotein. *Biochim. Biophys. Acta*, **2012**, 1820, 1664-1670.
- Kaminski, L.; Lurie-Weinberger, M. N.; Allers, T.; Gophna, U.; Eichler, J.: Phylogenetic- and genome-derived insight into the evolution of N-glycosylation in Archaea. *Mol. Phyl. Evol.*, **2013**, 68, 327-339.
- Kandt, C.; Ash W. L.; Tieleman D. P.: Setting up and running molecular dynamics simulations of membrane proteins. *Methods.*, **2007**, 41, 475-488.
- Kelleher, D. J.; Gilmore, R.: An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology*, **2006**, 16, 47R–62R.
- Kelly, J.; Jarrell, H.; Millar, L.; Tessier, L.; Fiori, L. M.; Lau, P. C.; Allan, B.; Szymanski, C. M.: Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J. Bacteriol.* **2006**, 188, 2427–2434.
- Ko, J.; Lee, D.; Park, H.; Coutsiyas, E. A.; Lee, J.; Seok, C.: The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res.*, **2011**, 39, W210-W214.
- Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M.: Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.*, **2006**, 25, 1957–1966.
- Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M.: N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science*, **2006**, 314, 1148–1150.
- Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G.: Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.*, **2008**, 77, 521–555.
- Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G.: ClustalW and ClustalX version 2. *Bioinformatics*, **2007**, 23, 2947–2948.
- Larkin, A.; Imperiali, B.: The expanding horizons of asparagine-linked glycosylation. *Biochemistry*, **2011**, 50, 4411-4426.

- Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M.: PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **1993**, *26*, 283-291.
- Leach, A. R.: Molecular Modelling: Principles and Applications. 2nd Ed., **2001**, Longman, Cingapura.
- Lee, J.; Lee, D.; Park, H.; Coutsias, E. A.; Seok, C.: Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins.*, **2010**, *78*, 3428-3436.
- Lehle, L.; Strahl, S.; and Tanner, W.: Protein glycosylation, conserved from yeast to man: A model organism helps elucidate congenital human diseases. *Angew. Chem. Int.*, **2006**, *Ed. 45*, 6802–6818.
- Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P.; X-ray structure of a bacterial oligosaccharyltransferase. *Nature*, **2011**, *474*, 350–355.
- Lole, K. S.; Bollinger, R. C.; Paranjape, R. S.; Gadkari, D.; Kulkarni, S. S.; Novak, N. G.; Ingersoll, R.; Sheppard, H. W.; Ray, S. C.: Full-length human immunodeficiency virus type 1 genomes from subtype C infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* **1999**, *73*, 152–160
- Maita, N.; Nyirenda, J.; Igura, M.; Kamishikiryo, J.; Kohda, D.: Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.*, **2010**, *285*, 4941–4950.
- Mally, M.; Fontana, C.; Leibundgut-Landmann, S.; Laacisse, L.; Fan, Y. Y.; Widmalm, G.; Aebi, M.: Glycoengineering of host mimicking type-2 LacNAc polymers and Lewis X antigens on bacterial cell surfaces. *Mol. Microbiol.*, **2013**, *87*, 112–131.
- Matsumoto, S.; Igura, M.; Nyirenda, J.; Matsumoto, M.; Yuzawa, S.; Noda, N.; Inagaki, F.; Kohda, D.: Crystal Structure of the C-Terminal Globular Domain of Oligosaccharyltransferase from *Archaeoglobus fulgidus* at 1.75 Å Resolution. *Biochemistry*, **2012**, *51*, 4157-4166.
- Matsumoto, S.; Shimada, A.; Kohda, D.: Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Struct. Biol.* **2013**, *13*, 11-19.

- Mor, A.; Ziv, G.; Levy, Y.: Simulations of proteins with inhomogeneous degrees of freedom: The effect of thermostats. *J. Comput. Chem.*, **2008**, *29*, 1992-1998.
- Neuberger, A: Carbohydrates in protein: the carbohydrate component of crystalline egg albumin. *Biochem. J.*, **1938**, *32*, 1435-1451.
- Nosé, S.; Klein, M. L.: Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **1983**, *50*, 1055-1076.
- Nosé, S.: A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **1984**, *52*, 255-268.
- Nothaft, H.; Szymanski, C. M.: Protein glycosylation in bacteria: Sweeter than ever. *Nat. Rev. Microbiol.*, **2010**, *8*, 765–778.
- Nothaft, H.; Scott, N. E.; Vinogradov, E.; Liu, X.; Hu, R.; Beadle, B.; Fodor, C.; Miller, W. G.; Li, J.; Cordwell, S. J.; Szymanski, C. M.: Diversity in the protein N-glycosylation pathways within the *Campylobacter* genus. *Mol. Cell Proteomics*, **2012**, *11*, 1203–1219.
- Nothaft, H.; Szymanski, C. M.: Bacterial protein N-glycosylation: New perspectives and applications. *J. Biol. Chem.*, **2013**, *288*, 6912-6920.
- Olivier, N. B.; Chen, M. M.; Behr, J. R.; Imperiali, B.: In vitro biosynthesis of UDP-N,N'-diacetylglucosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry*, **2006**, *45*, 13659–13669.
- Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F.: A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, **2004**, *25*, 1656-1676.
- Parrinello, M.; Rahman, A.: Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **1981**, *52*, 7182-7190.
- Paton, A. W.; Morona, R.; Paton, J. C.: Neutralization of shiga toxins Stx1, Stx2c, and Stx2e by recombinant bacteria expressing mimics of globotriose and globotetraose. *Infect. Immun.*, **2001**, *69*, 1967–1970.
- Paton, A. W.; Jennings, M. P.; Morona, R.; Wang, H.; Focareta, A.; Roddam, L. F.; Paton, J. C.: Recombinant probiotics for treatment and prevention of enterotoxigenic *Escherichia coli* diarrhea. *Gastroenterology*, **2005**, *128*, 1219–1228.

- Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D.: GROMOS 53A6<sub>GLYC</sub>, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theo. Comp.*, **2012**, *8*, 4681-4690.
- Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E.: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **2013**, *29*, 845-854.
- Rambaut, A.: FigTree. **2012** [<http://tree.bio.ed.ac.uk/software/figtree/>].
- Roth, J.: Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chem. Rev.*, **2002**, *102*, 285–304.
- Sayers, E. W.; Barrett, T.; Benson, D. A.; Bolton, E.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; Dicuccio, M.; Federhen, S.; Feolo, M.; Fingerman, I. M.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Krasnov, S.; Landsman, D.; Lipman, D. J.; Lu, Z.; Madden, T. L.; Madej, T.; Maglott, D. R.; Marchler-Bauer, A.; Miller, V.; Karsch-Mizrachi, I.; Ostell, J.; Panchenko, A.; Phan, L.; Pruitt, K. D.; Schuler, G. D.; Schriml, L. M.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Suzek, T. O.; Tatusov, R.; Tatusova, T. A.; Wagner, L.; Yaschenko, E.: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **2012**, *40*, D13–D25.
- Schachter, H.; Freeze, H. H.: Glycosylation diseases: Quo vadis?. *Biochim. Biophys. Acta, Mol. Basis Dis.*, **2009**, *1792*, 925-930.
- Schoenhofen, I. C.; McNally, D. J.; Vinogradov, E.; Whitfield, D.; Young, N. M.; Dick, S.; Wakarchuk, W. W.; Brisson, J.-R.; Logan, S. M.: Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: Enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.*, **2006**, *281*, 723–732.
- Sethuraman, N.; Stadheim, T.A.: Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.*, **2006**, *17*, 341–346.
- Spiro, R. G.: Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptides bonds. *Glycobiology*, **2002**, *12*, 43R-56R.

- Schaftenaar, G.; Noordik, J. H.: MOLDEN: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 123-134.
- Schüttelkopf, A. W.; van Aalten, D. M. F.: PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.*, **2004**, *D60*, 1355-1363.
- Schwarz, F.; Lizak, C.; Fan, Y.-Y.; Fleurkens, S.; Kowarik, M.; Aebi, M.: Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology*, **2011**, *21*, 45–54.
- Stepper, J.; Shastri, S.; Loo, T. S.; Preston, J. C.; Novak, P.; Man, P.; Moore, C. H.; Havlíček, V.; Patchett, M. L.; Norris, G. E: Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.*, **2011**, *585*, 645-50.
- Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P.: Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.*, 1999, *32*, 1022–1030.
- Tabish, S.; Raza, A.; Nasir, A.; Zafar, S.; Bokhari, H.: Analysis of glycosylation motifs and glycosyltransferases in Bacteria and Archaea. *Bioinformatics*, **2011**, *6*, 191-195.
- Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S.: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.*, **2011**, *28*, 2731–2739.
- Troutman, J. M.; Imperiali, B.: *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry*, **2009**, *48*, 2807–2816.
- Valderrama-Rincon, J.D.; Fisher, A. C.; Merritt, J. H.; Fan, Y. Y.; Reading, C. A.; Chhiba, K.; Heiss, C.; Azadi, P.; Aebi, M.; DeLisa, M. P.: An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.*, **2012**, *8*, 434–436.
- van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der

- Vegt, N. F. A.; Yu, H. B.: Biomolecular modeling: goals, problems, perspectives. *Angew. Chem. Int. Ed. Engl.*, **2006**, *43*, 4064–4092.
- Varki, A.: Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, **1993**, *3*, 97-130.
- Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M.: N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*, **2002**, *298*, 1790–1793.
- Wacker, M.; Feldman, M. F.; Callewaert, N.; Kowarik, M.; Clarke, B. R.; Pohl, N. L.; Hernandez, M.; Vines, E. D.; Valvano, M. A.; Whitfield, C.; Aebi, M.: Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems *Proc. Natl. Acad. Sci.*, **2006**, *103*, 7088–7093.
- Weerapana, E.; Glover, K. J.; Chen, M. M.; Imperiali, B.: Investigating bacterial N-linked glycosylation: Synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.*, **2005**, *127*, 13766–13767.
- Whelan, S.; Goldman, N.: A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.*, **2001**, *18*, 691–699.
- Yan, Q.; Lennarz, W.J.: Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.*, **2002**, *277*, 47692–47700.
- Yavuz, E.; Maffioli, C.; Ilg, K.; Aebi, M.; Priem, B.: Glycomimicry: display of fucosylation on the lipooligosaccharide of recombinant *Escherichia coli* K12. *Glycoconj. J.*, **2011**, *28*, 39–47.
- Yurist-Doutsch, S.; Abu-Qarn, M.; Battaglia, F.; Morris, H. R.; Hitchen, P. G.; Dell, A.; Eichler, J.: aglF, aglG and aglI, novel members of a gene cluster involved in the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. *Mol. Microbiol.*, **2008**, *69*, 1234–1245.

## 9. APÊNDICES

Trabalhos publicados durante o desenvolvimento do mestrado (Apêndices A e B) e artigo finalizado (Apêndice C) durante esse período, que foi enviado para obtenção do crédito de redação científica e que será submetido à revista *Langmuir*.

Apêndice A: Artigo publicado “Unrestrained Conformational Characterization of *Stenocereus eruca* Saponins in Aqueous and Nonaqueous Solvents”

Apêndice B: Artigo aceito para publicação “Improving the thrombin inhibitory activity of Glycyrrhizin, a triterpenic saponin, through a molecular simplification of the carbohydrate moiety”.

Apêndice C: Artigo em preparação para submissão “Atomic Model and Micelle Dynamics of QS-21 Saponin”.

## 9.1 APÊNDICE A

Pedebos, C; Pol-Fachin, L.; Verli, H. *J. Nat. Prod.* **2012**, *75*, 1196-1200.

# Unrestrained Conformational Characterization of *Stenocereus eruca* Saponins in Aqueous and Non- aqueous Solvents

Conrado Pedebos, Laercio Pol-Fachin, and Hugo Verli\*

## RESUMO

Saponins are secondary metabolites that have a plethora of biological activities. However, the absence of knowledge of their 3D structures is a major drawback for structural-based strategies in medicinal chemistry. To address this problem, the current work presents structural models of *Stenocereus eruca* saponins, named erucasaponin A and stellatoside B. These compounds were constructed based on a combination of unrestrained molecular dynamic (MD) simulations and NOESY data, in both pyridine and water. The models obtained in this way offer a robust description of the saponin dynamics in solution and support the use of submicrosecond MD simulations in describing and predicting glycoconjugate conformations.

## 9.2 APÊNDICE B

De Paula, F. T. *et al.* Chem. Biol. Drug. Des. "Accepted Article", doi: 10.1111/cbdd.12204

# Improving the anti-thrombin activity of Glycyrrhizin, a triterpenic saponin, through a molecular simplification of the carbohydrate moiety

Fernando T. de Paula<sup>a</sup>, Petrina Q. Frauches<sup>a</sup>, Conrado Pedebos<sup>b</sup>, Markus Berger<sup>c</sup>, Simone C. B. Gnoatto<sup>d</sup>, Grace Gossmann<sup>d</sup>, Hugo Verli<sup>b</sup>, Jorge A. Guimarães<sup>c</sup> and Cedric S. Graebin<sup>a, \*</sup>

Glycyrrhizin, a saponin, and its aglicone Glycyrrhetic acid are natural products found in the Licorice (*Glycyrrhiza glabra* L.) root extract. This saponin is known for its *in vitro* and *in vivo* Thrombin inhibitory activity. The design and synthesis of five glycyrrhizin derivatives was carried out in order to improve the natural product activity. Compound 3b, a phthalic ester derivative of Glycyrrhizin presented a more pronounced Thrombin inhibition ( $IC_{50} = 114.4 \pm 1.3 \mu M$ ) than the saponin ( $IC_{50} = 235.7 \pm 1.4 \mu M$ ). Molecular docking simulations done in order to investigate the molecular interaction between 3b and the enzyme indicate that this product is, as previously determined for Glycyrrhizin, an allosteric Thrombin inhibitor.

## 9.2 APÊNDICE C

# Atomic Model and Micelle Dynamics of QS-21

## Saponin

Conrado Pedebos, Laercio Pol-Fachin, Cilaine Verônica Teixeira, Ramon Pons and Hugo Verli.

### RESUMO

QS-21 is a saponin extracted from *Quillaja saponaria* (Soapbark tree), previously described as possessing anti-glycaemic and antitumoral activities, besides reducing plasma cholesterol and acting in the cardiovascular system. This biomolecule has been largely investigated as a vaccine immunoadjuvant. However, its use is mainly limited due to the chemical instability of this compound, added by a significant variety in molecular composition and a low tolerance dose. Another important feature of this saponin is its propensity to form micelles in a concentration-dependent manner, whose assembly was associated with a decrease in the hydrolysis of the ester bond related to its acyl chain. Considering that the comprehension of the molecular behavior of QS-21, at the atomic level, and knowledge of its micellar behavior, may support future studies involving the development of immunopotentiators, the current work aims to characterize the conformation and the process of micelle formation, inherent to the QS-21 saponins, experimentally and through computational techniques. Therefore, molecular dynamics simulations were performed in systems containing different numbers of QS-21 molecules in aqueous solution, in order to observe and evaluate the spontaneous micelle formation. The applied methodology allowed the generation of micelles whose sizes were compared to data obtained by SAXS. Furthermore, the ester linkage between the fucose residue and the acyl chain showed less solvation when in micellar form, suggesting a reduction in hydrolysis, representing a first atomistic interpretation of previous experimental data. Additionally, this is the first micellar

characterization of saponin micelles by SAXS and first tridimensional model of a micelle constituted of saponins, contributing to the understanding of the molecular basis of these compounds.

## 10. CURRICULUM VITÆ

### I. Formação acadêmica

Graduação em Biomedicina pela Universidade Federal do Rio Grande do Sul, de 2007/1 a 2011/2. *Caracterização Conformacional de Saponinas*. Orientador: Prof. Hugo Verli.

Mestrado em andamento em Biologia Celular e Molecular pelo Programa de Pós-Graduação em Biologia Celular e Molecular pela Universidade Federal do Rio Grande do Sul, de 2012/1 a 2013/1. *Glicobiologia Estrutural e Evolução Molecular da PglB de Campylobacter lari*. Orientador: Hugo Verli.

Doutorado em andamento em Biologia Celular e Molecular pelo Programa de Pós-Graduação em Biologia Celular e Molecular pela Universidade Federal do Rio Grande do Sul, a partir de 2013/2. *Glicobiologia Estrutural de Elementos da Via de N-glicosilação*. Orientador: Hugo Verli.

### II. Prêmios recebidos

**XXII Salão de Iniciação Científica, 2010:** Destaque de sessão na área de Ciências da Saúde - Sessão: Fitoquímica 1, UFRGS. Caracterização Conformacional de Saponinas. Orientador: Hugo Verli.

### III. Trabalhos científicos apresentados em congressos

**Pedebos, C.;** Verli, H.: Structural Biology of the Oligosaccharyltransferase PglB from *Campylobacter lari*. XLII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2013, Foz do Iguaçu / PR.

Graebin, C. S.; Paula, F. T.; Frauches, P. Q.; **Pedebos, C.;** Bergher, M.; Verli, H.; Guimarães, J. A.: Síntese e atividade inibidora de trombina de derivados da Glicirrizina e do ácido glicirretínico. 35ª Reunião Anual da Sociedade Brasileira de Química, 2012, Águas de Lindóia / SP.

**Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H.: Atomic Model for Micelles Composed by the Saponin QS-21. XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012, Foz do Iguaçu / PR.

**Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H.: Structure and Conformation of QS-21 Saponin: Implications for Micelle Formation. XL Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2011, Foz do Iguaçu / PR. Programa e Resumos da XL Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2011.

**Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H.: Caracterização Conformacional da Saponina QS-21 e as Implicações na Formação Micelar. XXIII Salão de Iniciação Científica, 2011, Porto Alegre / RS. Livro de resumos - XXIII Salão de Iniciação Científica e XX Feira de Iniciação Científica, 2011.

**Pedebos, C.;** Pol-Fachin, L.; Verli, H.: Depiction of Saponins Conformational Ensemble. XXXIX Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2010, Foz do Iguaçu / PR. Programa e Resumos da XXXIX Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2010.

**Pedebos, C.;** Pol-Fachin, L.; Verli, H.: Caracterização Conformacional de Saponinas. XXII Salão de Iniciação Científica, 2010, Porto Alegre / RS. Livro de resumos - XXII Salão de Iniciação Científica e XIX Feira de Iniciação Científica, 2010.

**Pedebos, C.;** Pol-Fachin, L.; Verli, H.: Conformational Characterization of Saponins Conformation in Pyridine. XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2009, Águas de Lindóia / SP. Programa e Resumos da XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2009.

**Pedebos, C.;** Pol-Fachin, L.; Verli, H.: Caracterização Conformacional de Saponinas em Piridina. XXI Salão de Iniciação Científica, 2009, Porto Alegre / RS. Caracterização Conformacional de Saponinas em Piridina, 2009.

#### **IV. Publicações em periódicos especializados**

**Pedebos, C.;** Pol-Fachin, L.; Verli, H. **2012.** Unrestrained conformational characterization of *Stenocereus eruca* saponins in aqueous and nonaqueous solvents. *J. Nat. Prod.*, *75*, 1196-1200.

De Paula, F. T.; Frauches, P, Q.; **Pedebos, C.**; Berger, M.; Gnoatto, S. C.; Gossmann, G.; Verli, H.; Guimarães, J. A.; Graebin, C. S. **2013**. Improving the thrombin inhibitory activity of Glycyrrhizin, a triterpenic saponin, through a molecular simplification of the carbohydrate moiety. *Chem. Biol. Drug. Des.* "Accepted Article", doi: 10.1111/cbdd.12204.

## **V. Bolsa recebida**

Bolsista FAPERGS de 2009/1 a 2011/1.

Bolsista CNPq – GM do programa de mestrado em biologia molecular e celular pelo Centro de Biotecnologia da UFRGS de março de 2013 a julho de 2013.

Bolsista CAPES do programa de doutorado em biologia molecular e celular pelo Centro de Biotecnologia da UFRGS a partir de agosto de 2013.