

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA
CELULAR E MOLECULAR

Conrado Pedebos

Estrutura e Dinâmica de Oligossacariltransferases
Procarióticas

Porto Alegre

2017

Conrado Pedebos

Estrutura e Dinâmica de Oligossacariltransferases Procarióticas

Tese 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 Doutor em Biologia Celular e Molecular.

Orientador: Hugo Verli

Porto Alegre

2017

Pedebos, Conrado

Estrutura e Dinâmica de Oligossacariltransferases Procarióticas/ Conrado
Pedebos. – Porto Alegre, 2017-
167 f.

Orientador: Hugo Verli

Tese (Doutorado) – 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, 2017.

1, N-glicosilação. 2, Oligossacariltransferases 3, Dinâmica Molecular. 4,
GROMOS I. Verli, Hugo, orient. II. Título

Conrado Pedebos

Estrutura e Dinâmica de Oligossacariltransferases Procarióticas

Tese 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 Doutor em Biologia Celular e Molecular.

Banca Examinadora:

Prof. Hugo Verli
Orientador

Prof. Ivarne Luis dos Santos Tersariol
Departamento de Bioquímica - UNIFESP

Prof. José Cláudio Fonseca Moreira
Departamento de Bioquímica - UFRGS

Prof. Munir Salomão Skaf
Departamento de Físico-Química -
UNICAMP

Prof. Nelson Jurandi Rosa Fagundes
(Revisor e Suplente)
Departamento de Genética - UFRGS

Porto Alegre, 04 de Agosto

2017

Agradecimentos

Agradeço, primeiramente, ao Prof. Hugo Verli por ter me aceitado e me orientado durante esses nove anos de laboratório. Pelos mais diversos ensinamentos e conselhos, buscando inspirar e dar suporte sempre que preciso. Pela constante preocupação com o desenvolvimento acadêmico e intelectual de seus alunos.

À minha comissão de acompanhamento, Dr. Charley Staats e Dr. Diego Bonatto, pelo suporte durante o desenvolvimento do trabalho.

Aos membros da banca, por terem aceitado o convite.

À secretaria do PPGBCM, em especial aos profissionais Silvia Regina Centeno e Luciano Saucedo, por toda dedicação, atenção e presteza com que realizam seu trabalho, sempre provendo auxílio aos alunos.

À UFRGS, ao Centro de Biotecnologia e ao Programa de Pós-Graduação em Biologia Celular e Molecular pela excelência do ensino que permitiu o meu desenvolvimento acadêmico.

Aos meus queridos colegas, amigos e agregados do Grupo de Bioinformática Estrutural (GBE), tanto os antigos como os atuais. Com vocês aprendi muito mais do que apenas aquilo que está escrito nos livros. Sem dúvida, vocês tornaram essa jornada muito mais agradável e divertida, rendendo momentos que guardarei com carinho. Faço questão de citá-los nominalmente aqui: Pablito (pela amizade, parcerias e humor peculiar), Rodrigo (pela mente inspiradora e pelos curiosidades bizarras), Laércio (pela amizade, paciência e por ter sido meu guia na modelagem molecular), Carla (pela parceria e conversas durante o doutorado), Marcelo (pelas *jam sessions*), Elisa (por rir das minhas piadas e pelo humor duvidoso), Juliano (Stoooooooooop), Bianca (por defender o indefensável), Fábio (pela boa insanidade), Felipe (por ser um IC abusado) e Bruno (pela construção do *Science Man*). Guys, foi um privilégio ter dividido a bancada com vocês, de verdade!

Aos meus pais, por todo carinho, amor e apoio que sempre me deram nas decisões que tomei nessa caminhada. Por sempre terem se dedicado a mim, fosse priorizando o meu bem-estar, fosse priorizando o meu desenvolvimento intelectual e humano.

À Soraia da Cas, minha incrível namorada/parceira/esposa/melhor amiga/inspiração/companheira de vida e de viagens! Agradeço-te pelo amor constante, suporte imenso, confiança e carinho durante todos esses anos. Pelo incentivo e inspiração para

que eu busque a melhor versão de mim, sem deixar que eu esmorecesse. Por ensinar a usar a magia em horas de dificuldade. Essa tese também é um pouco tua.

*“All things keep on in everlasting motion,
Out of the infinite come the particles,
Speeding above, below, in endless dance”*

Titus Lucretius Carus - On the Nature of Things

Resumo

A N-glicosilação é uma modificação distribuída de forma abundante nos domínios da vida, e é caracterizada pelo reconhecimento de uma assinatura N-X-S/T nas proteínas-alvo. O último passo dessa via biossintética é a transferência em bloco da cadeia glicídica para a asparagina da assinatura de N-glicosilação, sendo realizada pelas oligossacariltransferases (OSTs) PglB (Bacteria), AglB (Archaea) e STT3 (Eukarya). Dados cristalográficos para essas enzimas identificaram suas divisões em subunidades estruturais, como o núcleo central (CC), a sequência inserida (IS) e a periferia (P1 e P2). A unidade CC é fundamental para a interação com o substrato e a catálise, enquanto hipóteses da literatura sugerem que as outras unidades são importantes para a estabilidade estrutural das OSTs. Em decorrência do emprego desses sistemas enzimáticos em glicoengenharia de proteínas e no desenvolvimento de vacinas, a compreensão da origem estrutural para suas propriedades catalíticas e para a seletividade por substratos pode contribuir diretamente para o desenvolvimento de novas aplicações tecnológicas. Nesse âmbito, a presente tese avalia aspectos estruturais e conformacionais das OSTs, assim como a potencial função de suas unidades estruturais. Para tanto, empregamos a técnica de dinâmica molecular nas duas OSTs caracterizadas até o momento: a PglB de *Campylobacter lari* e a AglB de *Archaeoglobus fulgidus*. Desta forma, observamos a plasticidade conformacional do sítio ativo da PglB, com variações em elementos importantes, como a alça externa 5(EL5) e os resíduos catalíticos, sob influência da complexação dos substratos. Verificamos ainda que interações com as N-acetilações da glicana podem influenciar na seletividade desse substrato. Quanto à AglB, obtivemos evidências estruturais para propostas prévias da literatura de um papel termoestabilizador das unidades IS e P1 em relação à unidade CC. Nesse sentido, a presente tese esclareceu importantes aspectos estruturais e funcionais de OSTs, oferecendo suporte para o planejamento de novos experimentos e de aplicações tecnológicas. As novas informações sobre o mecanismo de seletividade da PglB podem auxiliar no desenvolvimento de glicoproteínas portando monossacarídeos específicos, potencialmente úteis para a produção de vacinas e de biomoléculas com uso terapêutico.

Palavras-chaves: Oligossacariltransferases. N-glicosilação. Dinâmica Molecular.

Abstract

N-glycosylation is a modification broadly distributed in all domains of life. It is characterized by a sequence motif N-X-S/T in proteins that portray this glycan chain. The last step of the oligosaccharide biosynthetic pathway is the *en bloc* transfer of the glycan chain to the Asn from the motif. This reaction is performed by enzymes named oligosaccharyltransferases PglB (Bacteria), AglB (Archaea), and STT3 (Eukarya). Previous crystallography data identified subdivisions in the OST domains, such as the central core (CC), the inserted sequence (IS), and the peripheries (P1 and P2). CC unit is essential to the interaction with substrates and catalysis. Hypothesis from the literature suggests that the other units are important for structural stability of OSTs. The N-glycosylation systems are being employed in studies regarding glycoengineering, and vaccine development, therefore, the comprehension of structural aspects that explains the catalytic properties, as well as substrate selectivity of OSTs, may contribute directly to the development of novel technological applications. In this context, this thesis evaluates the structural behavior of OSTs which involves the dynamics of conformational states, as well as the impact caused by the deletion of structural units, aiming to infer the function of these regions. To achieve these goals, molecular dynamics simulations were performed using both OSTs characterized until now: PglB from *Campylobacter lari* and AglB from *Archaeoglobus fulgidus*. Analyzing data generated by PglB simulations, it was possible to observe the conformational plasticity of the PglB active site, with variation of important elements, such as the external loop 5 (EL5) and the catalytic residues, when in the presence of different substrates. Regarding AglB data, we obtained structural evidences for the previously proposed thermostabilizing role displayed by IS and P1 units. In this manner, the present thesis clarified important structural and functional aspects of OSTs, offering support for new experiments and for technological applications. The new information about PglB selectivity mechanism could provide a basis for the development of glycoproteins possessing specific monosaccharides, potentially useful for the production of vaccines and therapeutical biomolecules.

Keywords: Oligosaccharyltransferase. N-glycosylation. Molecular Dynamics.

Lista de ilustrações

Figura 1 – Tipos de glicosilação	19
Figura 2 – Vias de N-glicosilação em Bacteria e em Eukarya	21
Figura 3 – Via de N-glicosilação em <i>Archaea</i> proposto por Eichler	25
Figura 4 – Organização Estrutural de OSTs	28
Figura 5 – Estrutura Parcial de <i>P. furiosus</i>	30
Figura 6 – Estruturas parciais de <i>A. fulgidus</i> e <i>C. jejuni</i>	30
Figura 7 – Estrutura completa da PglB de <i>C. lari</i> e seus principais motivos . .	31
Figura 8 – Estruturas completas da AglB de <i>A. fulgidus</i> em dois estados . . .	34
Figura 9 – Mecanismo de ação proposto para as OSTs	35
Figura 10 – Exemplos de aplicações biotecnológicas da glicoengenharia	39
Figura 11 – Estruturas tridimensionais obtidas por diferentes métodos	41
Figura 12 – Parâmetros que descrevem a forma funcional de um campo de força	44
Figura 13 – Esquema de modelagem do LLO de <i>C. lari</i>	58
Figura 14 – Comparação entre potenciais que descrevem o diedro estudado do LLO110	
Figura 15 – Distribuição dos componentes do sistema do LLO ao longo do eixo Z	111
Figura 16 – Complexo ternário formado pela PglB e seus substratos	111
Figura 17 – RMSD total e separado por domínios da PglB	112
Figura 18 – Estrutura de EL5 na simulação da PglB no estado pré-catálise . . .	114
Figura 19 – Distância da Tyr293 de EL5 em relação à N-acetilação em C4 da Bac115	
Figura 20 – Distâncias dos resíduos catalíticos para o metal divalente e a Asn .	116
Figura 21 – Distâncias essenciais para a reação catalítica da PglB durante a DM	117
Figura 22 – Análise de componentes principais da PglB em seu estado pré-catálise	118
Figura 23 – Análise de componentes principais da PglB em seu estado pós-catálise	120
Figura 24 – Análise de componentes principais da PglB em seu dois estados para um mesmo subespaço	121
Figura 25 – <i>Porcupine plot</i> diferenciando o estado inicial e a conformação de menor energia identificada por ACP do sistema pré-catálise	123
Figura 26 – <i>Porcupine plot</i> diferenciando o estado inicial do pós-catálise e a conformação de menor energia identificada por ACP para esse mesmo sistema	124
Figura 27 – Filogenia de OSTs realizada durante essa tese no método de Máxima Verossimilhança	128
Figura 28 – Filogenia de OSTs realizada pelo método de Máxima Verossimilhança por Lombard e comparação de estruturas baseadas na árvore filogenética	129

Figura 29 – Análises de estabilidade estrutural das PglBs quiméricas	132
Figura 30 – Ciclo de reação das OSTs	134

Lista de abreviaturas e siglas

DM	Dinâmica Molecular
OST	Oligossacariltransferase
TM	Domínio Transmembrana de OSTs
PP	Domínio Periplasmático de OSTs (ou C-terminal em Archaea e Eukarya)
CC	unidade estrutural denominada núcleo central (<i>central core</i>)
EL	Alça externa (<i>External Loop</i>)
IS	unidade estrutural denominada sequencia inserida (<i>inserted sequence</i>)
P1/P2	unidade estrutural denominada periferia 1 ou 2 (<i>periphery</i>)
MM	Mecânica Molecular
QM	Mecânica Quântica (<i>Quantum mechanics</i>)
NPT	<i>Ensemble</i> Isotérmico-Isobárico (número de partículas, pressão e temperatura constantes)
NVT	<i>Ensemble</i> Canônico (número de partículas, volume e temperatura constantes)
PDB	Banco de dados de estruturas proteicas (<i>Protein Data Bank</i>)
PME	<i>Particle Mesh Ewald</i>
CDG	Doenças Congênicas da Glicosilação (<i>Congenital disorders of glycosylation</i>)
GPI	Glicosilfosfatidilinositol
LLO	Oligossacarídeo ligado a lipídio (<i>Lipid-linked oligosaccharide</i>)
GTase	Glicosiltransferases
RMN	Ressonância Magnética Nuclear
LPS	Lipopolissacarídeo

RMSD	Raiz do desvio quadrático médio (<i>Root mean square deviation</i>)
RMS	Desvio quadrático médio (<i>Root mean square</i>)
DolPP	Dolicol-pirofosfato
DolP	Dolicol-monofosfato
UndPP	Undecaprenila-pirofosfato
THG	Transferência Horizontal de Genes
NOE	<i>Nuclear overhauser effect</i>
Glc	Glicose
GlcNAc	N-acetilglicosamina
GalNAc	N-acetilgalactosamina
Bac	Bacilosamina
PCA	Análise de Componentes Principais
DFT	Teoria do funcional da densidade
POPE	1-palmitoil-2-oleoil-sn-glicero-3-fosfoetanolamina

Lista de símbolos

ε	Constante dielétrica
K	Kelvin
β	Beta
α	Alfa
γ	Gama
Δ	Deleção
Å	Ângstrom
ρ	Densidade
ΔH_{vap}	Entalpia de Vaporização
J	Joule
μ	Micro
Da	Dalton

Sumário

1	INTRODUÇÃO	16
1.1	Glicobiologia e glicosilação	16
1.2	N-glicosilação	18
1.2.1	N-glicosilação em Bacteria	22
1.2.2	N-glicosilação em Archaea	24
1.3	Aspectos estruturais das OSTs	27
1.3.1	Anatomia de uma OST	27
1.3.2	Um histórico de estruturas parciais	29
1.3.3	A primeira estrutura completa: PglB	29
1.3.4	A segunda estrutura completa: AglB	32
1.3.5	O sítio catalítico e o mecanismo de ação das OSTs	33
1.4	Aplicações biotecnológicas da via de N-glicosilação e a glicoen- genharia de proteínas	36
1.5	Elucidação da estrutura e da conformação de biomoléculas	38
1.5.1	Validação das simulações por DM	44
1.5.2	Aplicação de abordagens computacionais para o estudo de OSTs	45
2	JUSTIFICATIVA	47
3	OBJETIVOS	48
4	PROCEDIMENTOS METODOLÓGICOS	49
4.1	Construção dos sistemas	50
4.2	Atracamento molecular	52
4.3	Inserção em membrana	53
4.4	Dinâmica molecular	53
4.5	Metadinâmica	56
4.6	Cálculos de mecânica quântica para a coordenação da PglB	56
4.7	Parametrização do LLO	56
4.8	Reconstrução filogenética	57
4.9	Análises	59
4.9.1	RMSD	59
4.9.2	RMSF	60
4.9.3	PCA	60

5	RESULTADOS	61
5.1	Capítulo I	62
5.2	Capítulo II	76
5.3	Capítulo III	109
5.3.1	Preâmbulo	109
5.3.2	Validação do modelo obtido para o LLO	109
5.3.3	Dinâmica global dos domínios e unidades estruturais	110
5.3.4	Coordenação da rede de interações no estado pré-catalítico	114
5.3.5	Caracterização do comportamento estrutural e conformacional da PglB	116
5.3.6	Análise de estados conformacionais da PglB	119
6	DISCUSSÃO GERAL	125
7	CONCLUSÕES	137
8	PERSPECTIVAS	138
	REFERÊNCIAS	139
	ANEXOS	156
	ANEXO A – ABP1	157
	ANEXO B – DDB1 E DDB2	159
	CURRICULUM VITÆ	161

1 Introdução

“Carbohydrates is all they groove”
Frank Zappa - Conehead

1.1 Glicobiologia e glicosilação

A glicobiologia pode ser definida, de forma bastante abrangente, como o estudo da estrutura, da biologia, da biossíntese e da evolução dos carboidratos (glicanas, açúcares, ou cadeias sacarídicas) distribuídos na natureza [1]. Até a primeira metade do século XX acreditava-se que os carboidratos serviam basicamente como fonte de energia, sem apresentarem funções biológicas distintas. Assim, durante a fase inicial da revolução da biologia molecular, nas décadas de 60 e 70, os estudos relativos a essa classe de biomoléculas encontravam-se muito pouco desenvolvidos em comparação às outras grandes classes moleculares. Contribuíram para isso a grande complexidade estrutural e a dificuldade na predição de sequências de glicanas, visto que essas não podiam ser obtidas diretamente do DNA. Somente a partir do final da década de 80, com a chegada de novas tecnologias empregadas na exploração da função e estrutura de glicanas, que o termo "glicobiologia" foi cunhado, reconhecendo a conjunção dos temas relacionados à química e à bioquímica de carboidratos com a biologia molecular e celular de glicanas e seus conjugados [2].

Essa classe de moléculas está espalhada de forma abundante na natureza, desempenhando funções que compreendem a geração de energia, a sinalização celular e a formação de componentes estruturais, dentre outras. Ainda, participam em diversos tipos de modificações co- ou pós-traducionais, produzindo os chamados glicoconjugados, como as glicoproteínas, os glicolipídeos e os proteoglicanos, ampliando a diversidade de composição das biomoléculas [1]. Entretanto, o paradigma central da biologia molecular, isto é, a informação biológica flui do DNA para o RNA e desse para as proteínas, não contempla o papel destas moléculas nos níveis de informação. Os glicoconjugados estão envolvidos em uma enorme variedade de eventos, como por exemplo interações célula-célula, bem como com a matriz extracelular e com outras moléculas, sendo essenciais para o funcionamento e o desenvolvimento de organismos multicelulares complexos. Além disso, estes conjugados vem sendo identificados como relevantes na patogenicidade de alguns organismos [3]. Assim, conforme postulou Varki [1], uma forma mais completa e adequada de descrever o paradigma central da biologia molecular seria com a inclusão

das glicanas e de suas combinações covalentes com outras macromoléculas.

O processo de formação de glicoconjugados é chamado de glicosilação. Essa é a modificação co- ou pós-traducional mais prevalente [4, 5] encontrada na natureza. Considera-se que cerca de dois terços das proteínas existentes no domínio Eukarya potencialmente apresentam essa característica [6]. A extensa variabilidade de carboidratos combinada à microheterogeneidade das glicanas e às diferentes formas de ligação entre eles possibilita a diversificação do proteoma, permitindo expandir a complexidade biológica dos genomas. A inclusão de uma glicana à estrutura de uma proteína pode impactar diretamente na sua função, localização, sinalização e adesão, dentre outros.

No início do século XX, Albert Neuberger obteve as primeiras evidências dessa característica em organismos a partir da detecção de monossacarídeos como manose e glicosamina [7]. Tal estudo serviu como base para a comprovação da existência de carboidratos ligados à estrutura de proteínas. Atualmente, uma enorme quantidade de funções tem sido relacionada à presença dessas moléculas na estrutura proteica, como na termodinâmica de enovelamento de proteínas [8], na sinalização celular, na localização intracelular, no reconhecimento intercelular, na resposta imunitária e na estabilidade contra proteólise [4]. Outro campo em que os processos de glicosilação foram identificados como relevantes é em patologias observadas na prática clínica [9], sendo um dos grandes exemplos os chamados Distúrbios Congênitos da Glicosilação (*Congenital Disorders of Glycosylation - CDG*), que já somam mais de 60 tipos diferentes descritos na literatura [10] e atingem as vias responsáveis pela biossíntese de glicoconjugados. Ainda, são crescentes os relatos de alterações nos padrões de glicosilação relacionados à iniciação, à progressão e à metástase de tumores, sendo a glicosilação considerada um marcador de câncer.

A glicosilação é um processo que pode ser de vários tipos (Figura 1). Até o momento, os já identificados são [5, 11, 12]:

- O-glicosilação, que envolve a adição de glicanas a um átomo de oxigênio de um resíduo de aminoácido, geralmente uma serina ou uma treonina;
- P-glicosilação, onde uma serina fosfatada terá o seu átomo de fósforo comoceptor da glicana;
- C-manosilação, que está relacionada à presença de carboidratos em um átomo de carbono do anel indólico de um resíduo de triptofano;
- S-glicosilação, que envolve a adição de monossacarídeos à cadeia lateral de um resíduo de cisteína, utilizando o átomo de enxofre ali presente;

- C-manosilação, que está relacionada à presença de carboidratos em um átomo de carbono do anel indólico de um resíduo de triptofano;
- Gliptação, descrita como a glicosilação de um lipídio chamado âncora de glicosil-fosfatidilinositol (GPI) e que pode ser adicionado ao ponto C-terminal de uma proteína para realizar o ancoramento a uma membrana celular;
- N-glicosilação, que é caracterizada por uma ligação do tipo β -glicosilamida de um oligossacarídeo ao átomo de nitrogênio da carboxamida presente na cadeia lateral de uma asparagina aceptora.

1.2 N-glicosilação

A N-glicosilação é o tipo de glicosilação de maior ocorrência já que, de todas as glicoproteínas, cerca de 90% corresponde a essa forma [6]. Essa modificação está distribuída pelos três domínios da vida [13, 14], tendo sido encontrada, primeiramente, em Eucariotos, novamente com estudos de Neuberger e colaboradores [15]. Por um longo período acreditou-se que a N-glicosilação fosse restrita ao domínio Eukarya. Tal noção começa a mudar em meados da década de 70, quando um estudo de Mescher & Strominger [16] demonstrou a presença de glicoproteínas na camada-S do organismo *Halobacterium salinarum*, abrindo caminho para estudos posteriores que identificaram N-glicosilação no flagelo de *H. salinarum* [17] e na camada-S de *Haloferax volcanii* [18], espécie encontrada no Mar Morto. Apenas algumas décadas depois esse processo foi descoberta em bactérias [19, 20], gerando uma lacuna de conhecimento quando comparado à amplamente estudada e conservada via eucariótica, principalmente por meio do modelo *Saccharomyces cerevisiae* [21]. Entretanto, nas últimas duas décadas ocorreu a massificação do sequenciamento genômico de procariotos, concomitantemente ao melhoramento das técnicas de espectrometria de massas, o que proporcionou a identificação de sistemas de N-glicosilação nas mais variadas espécies de bactérias e archaeas [22].

A via de N-glicosilação em bactérias que possui a maior quantidade de estudos, sendo, portanto, a mais bem compreendida até os dias de hoje, é a do microorganismo *Campylobacter jejuni* [3]. As vias existentes em archaeas, sobretudo devido à alta complexidade e variabilidade das cadeias de oligossacarídeos, ainda possuem pontos obscuros porém, aos poucos, o conhecimento a seu respeito vai sendo solidificado. De qualquer forma, os passos iniciais pertencentes a esse processo, os quais se referem à construção da cadeia oligossacarídica e sua consequente transferência para uma proteína

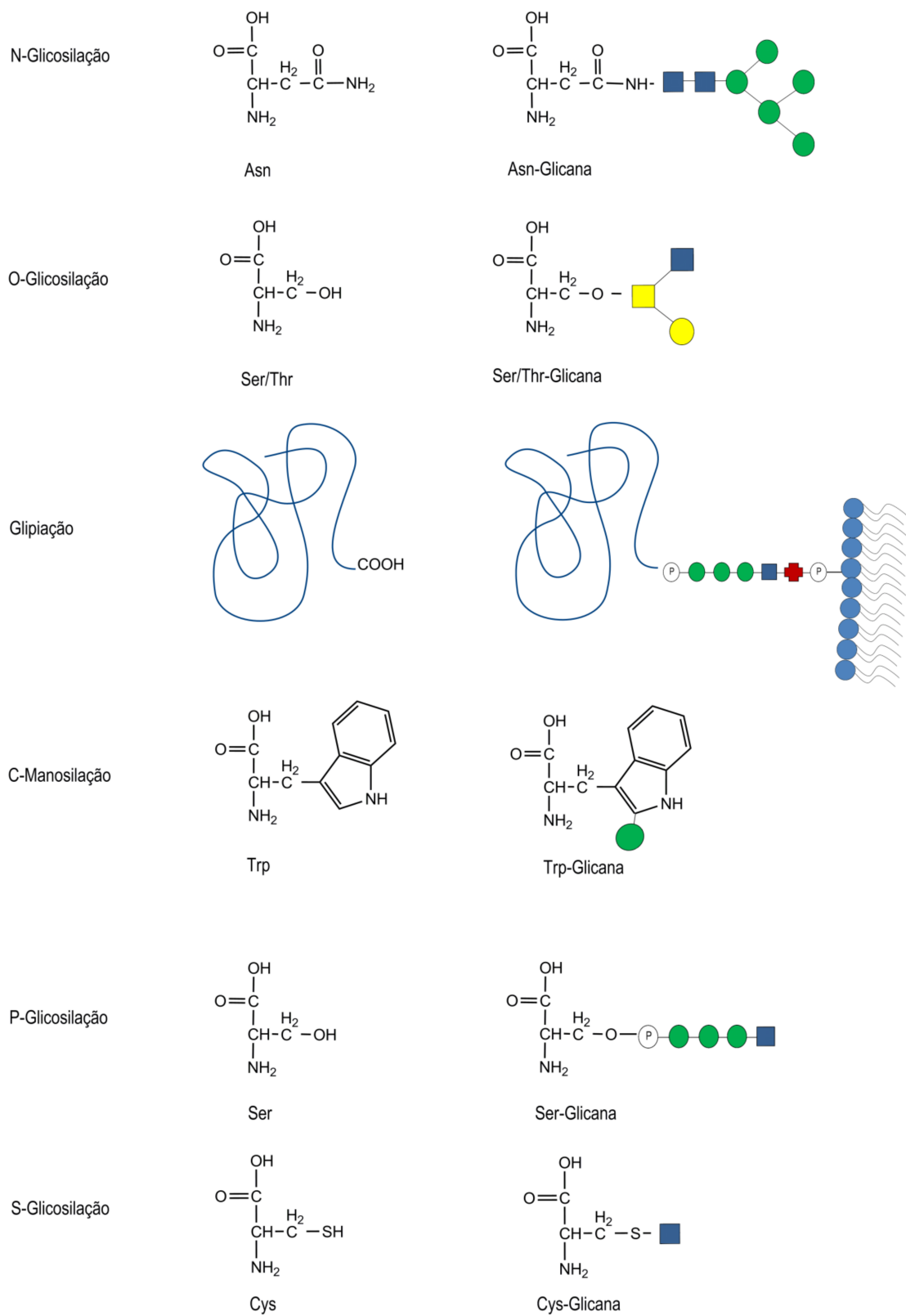


Figura 1 – Diferentes tipos de glicosilação com exemplos de átomos e resíduos em que carboidratos podem ser adicionados.

acceptora, ocorrem com um certo grau de conservação em todos os domínios da vida (Figura 2). A logística básica dessa via envolve a participação de diversas enzimas que irão atuar na biossíntese do oligossacarídeo, o qual é conectado e transportado via um carreador lipídico, formando um oligossacarídeo ligado a lipídio (em inglês, *Lipid-linked oligosaccharide* - LLO). Esse carreador pode ser em bactérias uma molécula de undecaprenila pirofosfatada (Und-PP) ou um dolicol, tanto mono como difosfatado em arqueas ou somente difosfatado em eucariotos. Quanto às enzimas da via [22] responsáveis pela montagem do substrato sacarídico, em eucariotos são as glicosiltransferases (GTases) da família *asparagine-linked glycosylation* (Alg), que adicionam os monossacarídeos um a um ao carreador. Similarmente, em arqueas as enzimas do tipo GTases da família *archaeal glycosylation* (Agl) estão envolvidas no processo, enquanto em bactérias as responsáveis são as da família *protein glycosylation* (Pgl). Essas enzimas utilizam apenas formas ativas dos monossacarídeos (também chamados de glicídeos nucleotídicos) como doadores para as reações de construção da cadeia oligossacarídica.

Em Bacteria, toda essa linha de montagem irá ocorrer com o carreador lipídico e seu açúcar voltados para o citoplasma [22]. Em Archaea, existem algumas exceções, onde ocorrem edições posteriores, enquanto em Eukarya a árvore glicídica e seu lipídio são translocados para o lúmen do retículo endoplasmático após a adição de apenas uma parte dos monossacarídeos. É importante ressaltar, ainda, que devido à microheterogeneidade dos carboidratos, faz-se necessário o uso de diferentes enzimas de uma mesma família que comportem os variados substratos dessa via, podendo inclusive ser únicas para cada espécie. Para que a transferência da glicana seja executada é essencial que o carreador lipídico esteja localizado no compartimento celular apropriado: em eucariotos, a adição da glicana à proteína ocorre 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 LLO esteja ao alcance das enzimas denominadas oligossacariltransferases (OST), a ação de transportadores do tipo ABC com função de flipase é imperativa, pois são eles que realizam a translocação desse carreador glicosilado a partir da realização do movimento de *flip*. No domínio Eukarya, a proteína Rft1 era considerada a responsável por essa função [23], contudo essa hipótese foi refutada posteriormente após novas investigações, nas quais foi descoberto que se trata apenas de uma coadjuvante neste processo [24], restando ainda incerto qual transportador desempenha esse papel. Enquanto em bactérias a enzima PglK é designada como a flipase envolvida [25], em archaeas, na via da espécie *Haloferox volcanii*, foi detectado que a enzima AglR pode desempenhar essa função [26], embora nesse domínio da vida ainda sejam necessários mais estudos que elucidem o transportador ou os transportadores existentes. Após a translocação

figuras/Figure2.pdf

Figura 2 – Vias de N-glicosilação em *Bacteria* (*C. jejuni*) e em Eukarya (*S. cerevisiae*) demonstrando o padrão sequencial em cada caso. Adaptado de Larkin & Imperiali [22].

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).

O passo final da via é a transferência da árvore glicídica para a proteína a ser glicosilada. Para tanto, é vital o papel desempenhado pelas OSTs, enzimas catalisadoras dessa reação. Cada enzima dessas pode variar amplamente em sua quantidade de resíduos de aminoácidos, mudando de espécie para espécie, e também em termos de características estruturais. As OSTs eucarióticas, via de regra, formam um complexo composto de 5 a 8 subunidades, incluindo a STT3 (*staurosporine and temperature sensitive 3*), que exerce a função catalítica [27,28]. Diversamente, as archaeas e bactérias possuem apenas uma subunidade, a AglB (*Archaeal glycosylation B*) [29] e a PglB (*Protein glycosylation B*) [30], respectivamente, com função idêntica a da STT3 de eucariotos. Uma ressalva importante é que alguns protistas também apresentam apenas uma subunidade, a catalítica STT3 [31]. Ainda, uma mesma espécie é capaz de possuir mais de uma isoforma de subunidade catalítica, dotadas inclusive de glicanas específicas e com baixas identidades de sequência entre si [32]. Com o processo de transferência executado, a proteína pode tanto manter a glicana na forma que foi transferida como sofrer alterações adicionais na sua cadeia oligossacarídica, promovidas no retículo endoplasmático e no Complexo de Golgi (eucariotos) ou realizadas por enzimas adicionais (archaea), processadoras da glicana [33,34]. No domínio Bacteria, esse tipo de modificação ainda não foi encontrado; todavia, em eucariotos a principal fonte de diversidade estrutural associada à carboidratos é advinda desse pós-processamento.

1.2.1 N-glicosilação em Bacteria

Dentre o domínio Bacteria, as espécies da classe Epsilonproteobacteria são aquelas que mais apresentam o *cluster* gênico que codifica o sistema de N-glicosilação [3,13]. Dentro dessa classe, o gênero mais estudado é o *Campylobacter*, destacando-se ainda os gêneros *Helicobacter* e *Wolinnella*. Alguns integrantes da classe Deltaproteobacteria, como o gênero *Desulfovibrio*, possuem seu sistema de N-glicosilação caracterizado [35]. Adicionalmente, um sistema bastante particular foi descrito na classe Gamaproteobacteria – a GTase HMW1C, cuja atividade é desempenhada no citoplasma ao glicosilar a adesina HMW1 [36]. Essa OST possivelmente evoluiu de forma independente, pois possui pouca similaridade estrutural com a PglB, além de estarem localizadas em compartimentos celulares diferentes (citoplasma e periplasma, respectivamente) [3].

A diversidade de estruturas sacarídicas N-ligadas nestes microorganismos é um dos temas que vem sendo constantemente estudados. O aspecto mais conservado 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 [20]. Portanto, faz sentido 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, todas modificando a posição 4 do monossacarídeo GlcNAc. Inicialmente a PglF, uma desidratase, retira o átomo de hidrogênio (H) do grupo hidroxila (OH) a partir de uma molécula de NAD⁺, que se tornará NADH. A aminotransferase PglE acrescenta o grupo amina na sequência e remove o grupo cetona ali presente [37]. Finalizando o processo, a PglD transfere a acetilação utilizando uma molécula de acetil-CoA [38], assim formando a UDP-Bac que possui um grupo pirofosfato em sua posição 1. A PglC é uma fosfoglicosiltransferase que age na transferência da Bac fosforilada na posição 1 (Bac-1-P) para o carreador lipídico (Und-P) [39]. As PglA e PglJ [40, 41] adicionam a primeira e a segunda GalNAc com ligações $-(1\rightarrow3)$ e $-(1\rightarrow4)$, respectivamente. A PglH é uma polimerase, sendo responsável por anexar os próximos três monossacarídeos de N-acetilgalactosamina (GalNAc) ligados de forma $-(1\rightarrow4)$ [42]. O passo final da montagem é o acréscimo do monossacarídeo que ramifica a glicana, a partir da ação da PglI que adiciona um resíduo de glicose (Glc) com uma ligação $-(1\rightarrow3)$ [41].

Após o processo realizado na linha de montagem do heptassacarídeo, a OST de bactérias, denominada PglB, irá catalisar a transferência da glicana. Para isso, a proteína acceptora da glicosilação necessita possuir uma assinatura em sua sequência. A assinatura padrão básica para N-glicosilação 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. Contudo, a PglB demanda uma assinatura estendida para obter maior eficiência [43], apesar desse requerimento não ser obrigatório para a ocorrência da modificação pós-traducional [44]. Em bactérias, o motivo é composto por uma extensão D/E-X-N-X-S/T, onde os dois resíduos de aminoácido extras 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, há uma proposição de que a região desse motivo precisa estar em uma alça flexível e exposta para que, assim, possa interagir com o sítio catalítico da PglB. Entretanto, estudos recentes desafiam essa noção [45], indicando que o processo pode ocorrer co-traducionalmente, com as proteínas apenas parcialmente enoveladas.

1.2.2 N-glicosilação em Archaea

A N-glicosilação em Archaea é considerada a modificação pós-traducional capaz de gerar a maior diversidade química dentre os três domínios da vida [44, 46]. Além dos tipos de monossacarídeos diferentes, possui diversas alterações em seus monossacarídeos, como metilação, sulfatação e a adição de treoninas. As análises de genomas realizadas demonstraram a existência de vias praticamente espécie-específicas [47]. Apesar dessa diversidade, os esforços direcionados à elucidação dessas vias no domínio Archaea tem sido limitados a uma diminuta quantidade de espécies, dentre elas *H. volcanii*, *Methanococcus maripaludis*, *Methanococcus voltae* e *Sulfolobus acidocaldarius*. Algumas iniciativas criadas por pesquisadores da área, como Eichler e colaboradores, envolveram a criação de um banco de dados, o *aglgenes* [48], que busca o monitoramento dos genes relacionados a essa via para diferentes espécies, bem como para a confecção de uma proposta de padronização na sua nomenclatura [49], contribuindo à identificação de características comuns entre eles.

Dentre as vias estudadas em Archaea, a de *H. volcanii*, do filo Euryarcheota, é tida como a mais bem compreendida [46, 50–53] (Figura 3), embora alguns elementos ainda sejam desconhecidos. A enzima que adiciona o primeiro monossacarídeo, uma hexose, ao carreador lipídico é a enzima AglJ [54], seguida pelas ações das GTases AglG, AglI e AglE, que irão transferir três unidades de ácidos hexurônicos [29]. Concomitantemente, as enzimas AglM, AglQ e AglP atuarão em funções auxiliares na construção dessa estrutura [55–57]. A primeira age como uma UDP-glicose desidrogenase na biossíntese das duas primeiras unidades de ácido hexurônico. A segunda é descrita com função de epimerase/isomerase e é importante para o aparecimento da terceira unidade de ácido hexurônico, sem ainda uma função específica conhecida. A terceira é uma S-adenosil-L-metionina-dependente metiltransferase, desempenhando uma metilação na posição quatro da terceira unidade de ácido hexurônico. A enzima que realiza a translocação desse tetrassacarídeo ligado ao dolicol-fosfato (DolP) ainda é desconhecida. Geralmente, a etapa derradeira do processo ocorre quando a oligossacariltransferase AglB catalisa a transferência em bloco da cadeia oligossacarídica para a proteína aceptora. Contudo, no caso da via de *H. volcanii*, ocorre um processo posterior à construção e à adição desse tetrassacarídeo: o acréscimo de um resíduo de manose que completará o pentassacarídeo de *H. volcanii*. Para tal fim, a enzima AglD irá realizar a transferência de um monossacarídeo de manose para outro DolP; a enzima AglR realiza a translocação desse LLO do citoplasma para o espaço entre a membrana e a parede celular; e a enzima AglS catalisa a transferência desse monossacarídeo para o tetrassacarídeo, completando assim a glicana nas proteínas desse organismo.

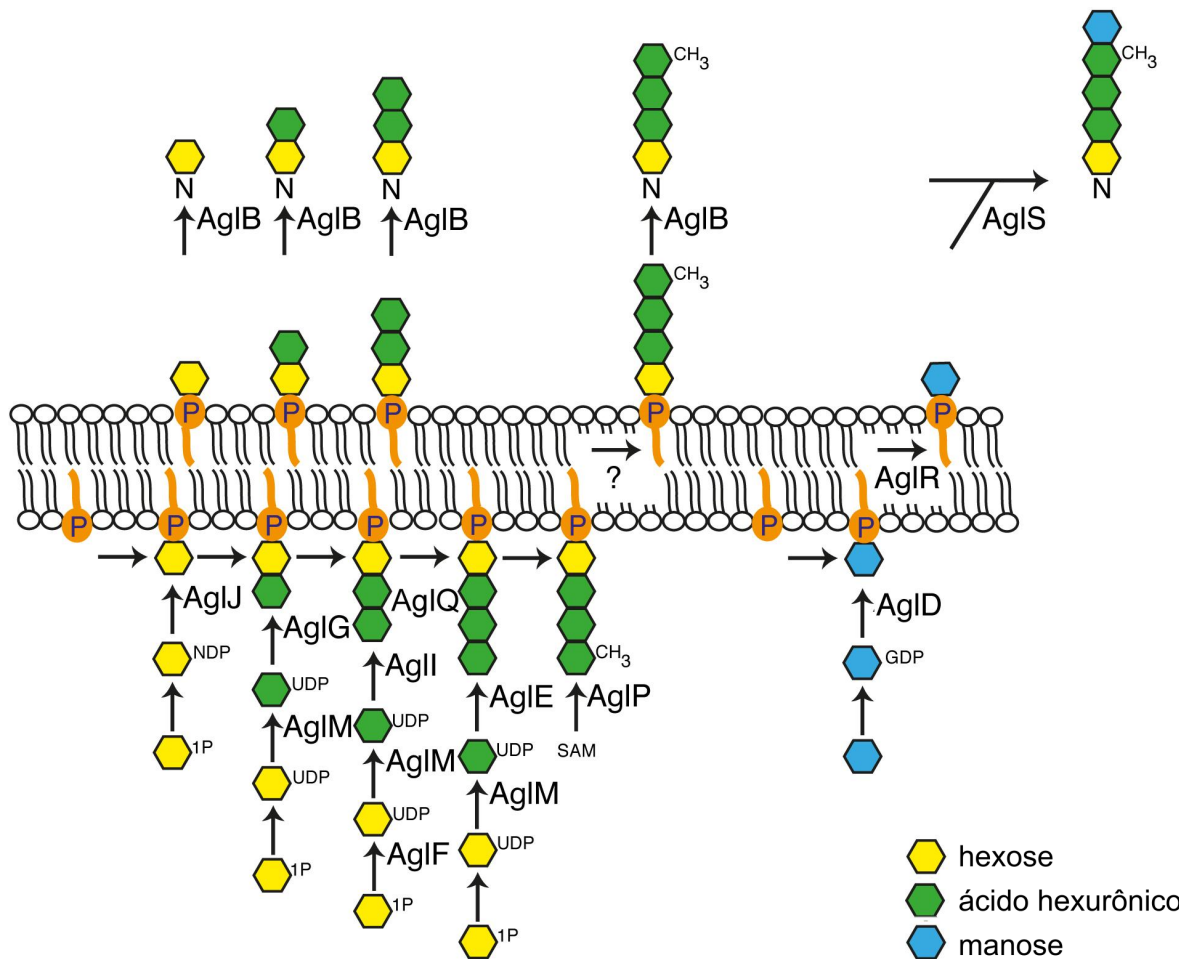


Figura 3 – Proposta de via de N-glicosilação em *Archaea* (*H. volcanii*) feita por Eichler e colaboradores [50]. As enzimas AglJ, AglG, AglI e AglE transferem uma hexose e três ácidos hexurônicos, respectivamente. Enquanto isso, a AglM, a AglQ e a AglP participam secundariamente na formação das unidades glicídicas. A AglB transfere o tetrassacarídeo para a proteína acceptora. Posteriormente, um resíduo de manose é adicionado ao tetrassacarídeo, em um processo que envolve a AglD (adição da manose ao carreador), a AglR (flipase) e a AglS (transferência para a glicana).

Nessa mesma espécie ocorre um fenômeno de regulação bastante particular: a partir da exposição a diferentes concentrações salinas, o padrão de glicosilação de uma glicoproteína é alterado, de modo que uma segunda glicana (estruturalmente diversa da primeira) é adicionada por uma segunda via de N-glicosilação [58]. Na presença de 1,75 M de NaCl (mas não 3,4), a glicoproteína da camada-S recebe um tetrassacarídeo (ao invés do penta) na Asn498 enquanto, anteriormente, a proteína era glicosilada na Asn13 e na Asn83. Esse segundo processo é controlado por genes de uma região distinta no genoma, compreendendo os genes HVO_2046 ao HVO_2061, inclusive utilizando uma oligossacariltransferase ainda desconhecida. Normalmente, em altas concentrações salinas essa modificação possui níveis baixíssimos, esparsamente detectados. Porém, surpreendentemente, após a deleção do *cluster* gênico relativo à primeira via, as proteínas da camada-S começam a receber a modificação na Asn498, mesmo com a concentração salina que antes não favorecia seu aparecimento, um fato que ainda carece de maiores estudos para sua compreensão e que evidencia que há algum grau de comunicação entre as vias.

Ao contrário do que foi visto para outras espécies do domínio Archaea (como a própria *H. volcanii*), a N-glicosilação é essencial em *Sulfolobus acidocaldarius*, pertencente ao filo Crenarcheota. A deleção do gene que codifica a AglB é incompatível com a vida [59], mostrando a importância dessa enzima para a sobrevivência dessa espécie. Isso pode ser justificado pela grande quantidade de sítios de N-glicosilação na glicoproteína da camada-S [60], tornando possível o enfrentamento das altas temperaturas e do baixo pH. A via biossintética ainda possui diversas lacunas na sua descrição, porém parece começar com a ação da AglH, que transfere um monossacarídeo de UDP-GlcNAc para o lipídio carreador dessa espécie, um dolicol pirofosfatado (DolPP) com cadeias hidrofóbicas de tamanho reduzido (35 a 40 carbonos). Os dois passos seguintes da via, a adição de mais uma GlcNAc e de uma manose, formando um trissacarídeo, ainda não possuem enzimas devidamente identificadas. No quarto passo, a Agl3 converte a UDP-glicose e sulfato de sódio em uma UDP-sulfoquinovose, que na sequência é transferida para o trissacarídeo por uma GTase desconhecida. Após esse passo, mais duas unidades são adicionadas à glicana: uma manose e uma glicose, sendo essa última ligada por ação da enzima Agl16. Postula-se que o LLO completo é translocado do citoplasma para o espaço externo por intermédio de uma flipase ainda desconhecida. O último passo é catalisado pela AglB, que irá transferir a glicana em bloco.

1.3 Aspectos estruturais das OSTs

1.3.1 Anatomia de uma OST

As OSTs já foram identificadas em diversos organismos espalhados pelos três domínios da vida. Entretanto, apesar de sua importância e de sua essencialidade, o grau de similaridade entre elas é bastante baixo. Grandes variações estruturais são observadas, tanto em quantidade de resíduos (com algumas enzimas apresentando cerca de 550 resíduos e outras até 1500), como em conservação do enovelamento e presença de subunidades. Acrescenta-se ainda o fato de que algumas espécies apresentam mais de uma oligossacariltransferase em seu genoma, por vezes com baixíssima identidade entre elas. Todavia, ao observar as estruturas caracterizadas até hoje (Figura 4), percebe-se que padrões emergem e é possível inferir alguma conservação estrutural entre elas [32]. O domínio transmembrana (TM) é formado em média por 13 hélices agrupadas, que são conectadas por alças intracelulares e externas. Dentre essas alças externas (em inglês, *external loop* - EL), sabe-se que algumas desempenham papéis fundamentais no mecanismo de catálise e ligação ao substrato. Na interface entre os domínios, coordenado por resíduos do domínio TM, está o íon metálico essencial (M^{2+}) para a catálise, responsável pela coordenação dos resíduos que formam o sítio catalítico e que permitirão a formação da rede de interações que permeia o mecanismo de ação das OSTs. Diferentes íons já foram observados exercendo este papel, tais como Mg^{2+} , Mn^{2+} , Zn^{2+} e Co^{2+} , com grande variação de atividade dependendo do metal de transição utilizado e da concentração aplicada no estudo [61].

O domínio PP ou C-terminal possui organizações que variam conforme cada espécie e pode ser subdividido em diferentes unidades estruturais [62, 63]. Aquele que parece ter presença obrigatória e possui atribuições essenciais nas OSTs é o núcleo central (em inglês, *central core* - CC), já que é a unidade que possui dois importantes motivos conservados WWDXGX e MI/DK cuja função é reconhecer o S/T da posição +2 na assinatura de N-glicosilação [64]. O enovelamento dessa região varia mas, em geral, é um misto α/β , podendo apresentar uma maior quantidade de α -hélices em eucariotos e uma maior quantidade de β -fitas nos procariotos já estudados. A unidade denominada sequência de inserção (em inglês, *insertion sequence* - IS) foi identificada até o momento apenas em Bactéria e em Archaea, e é assim denominada pois parece ter sido inserida dentro da sequência do domínio CC. Ela ocorre em Bactéria e em Archaea e, nas estruturas elucidadas até agora, essa unidade apresentou-se com estrutura secundária formada por β -folhas, demonstrando um padrão variável de enovelamento e tamanho. Inicialmente, foi sugerido que essa unidade possuía envolvimento no reconhecimento do substrato sacarídico [63] entretanto, como será mostrado mais adiante, a elucidação da

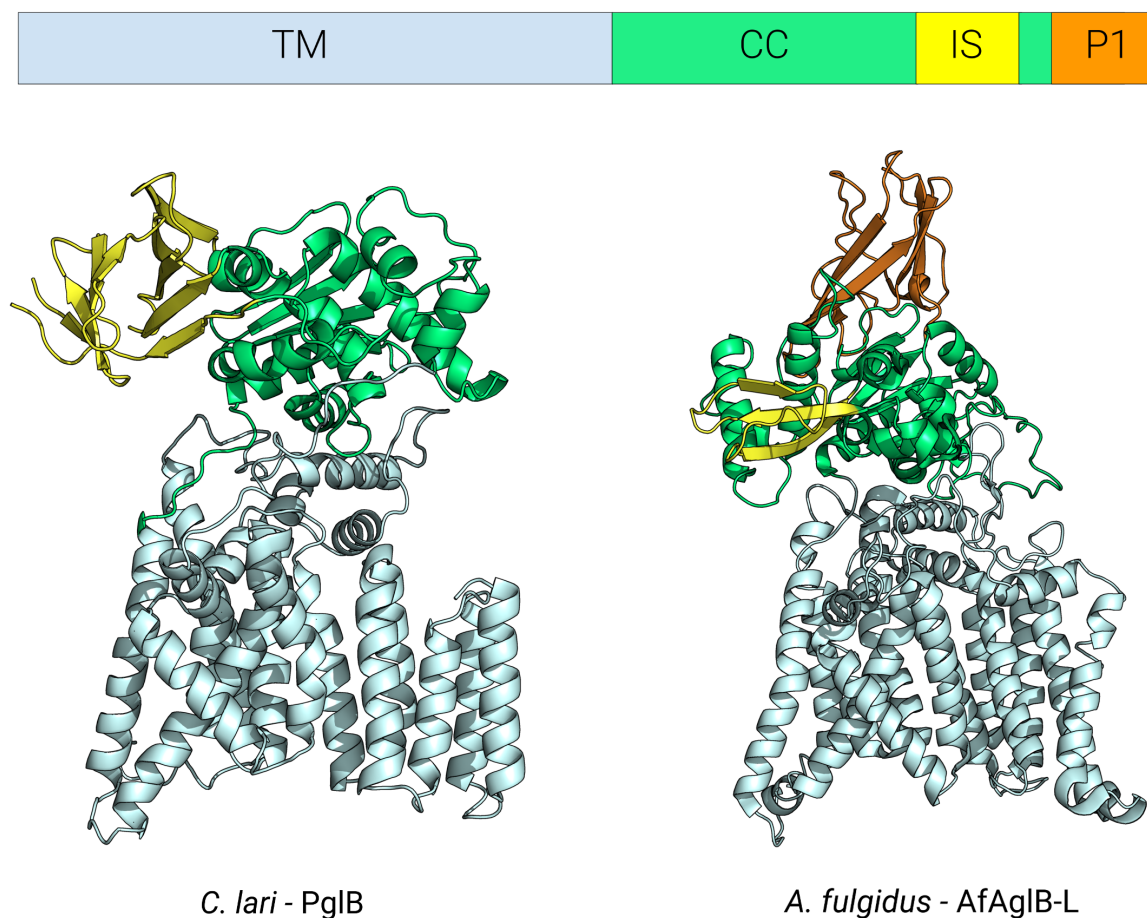


Figura 4 – Organizações estruturais já encontradas para OSTs completas demonstrando a variabilidade entre espécies. O domínio TM e as unidades estruturais estão coloridas de acordo com o código.

estrutura completa da PglB descartou essa hipótese [64]. O terceiro tipo de unidade estrutural já identificado inclui as periferias 1 e 2 (em inglês, *periphery* 1 e 2 - P1 e P2), que parece envolver a parte externa da unidade CC. Uma OST pode apresentar apenas P1 ou as duas estruturas, P1 e P2. Com exceção do CC, as outras unidades estruturais ainda não tem função conhecida. Ensaio de mutagênese verificaram que a unidade IS da AglB de *P. furiosus* é dispensável para catálise [65]. Todavia, o ensaio foi realizado com um pequeno peptídeo acceptor e com a enzima solubilizada (na ausência de membrana), condições que podem influenciar, por exemplo, na interação com proteínas maiores ou no acoplamento com proteínas de membrana, onde o polipeptídeo a ser glicosilado estará sendo translocado. As unidades P1 e P2 são sugeridas na literatura como possíveis fatores termoestabilizantes da estrutura das OSTs [66], já que muitas delas são encontradas em organismos hipertermófilos, hipótese ainda por ser confirmada.

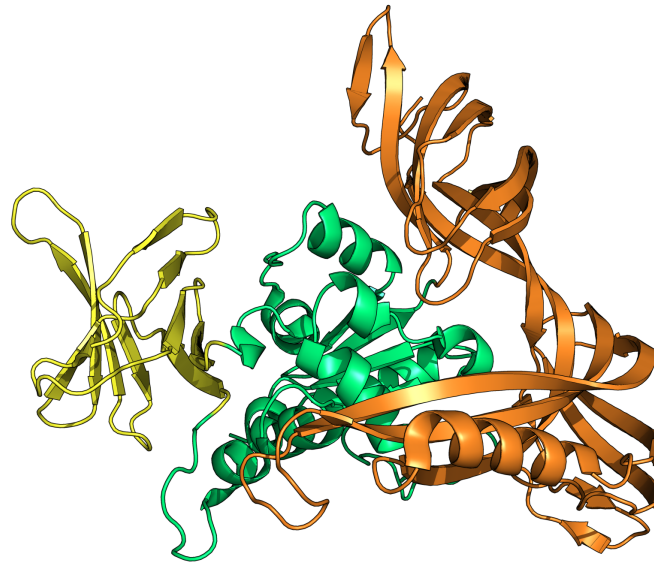
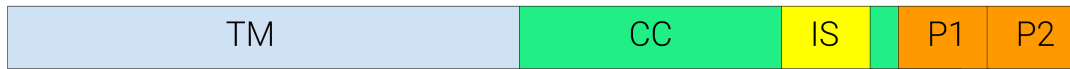
1.3.2 Um histórico de estruturas parciais

Conforme apontado anteriormente, um dos marcos do processo de N-glicosilação ocorre quando a glicana completa (ou que será processada posteriormente) ligada ao lipídio carreador é transferida por uma OST para o acceptor proteico, gerando as glicoproteínas. A função de uma proteína está diretamente ligada a sua estrutura, portanto a obtenção de dados estruturais de OSTs é uma área em constante desenvolvimento. Até 2011, nenhuma estrutura completa de OSTs havia sido obtida, restando apenas as elucidações de porções completas ou incompletas do denominado domínio periplasmático (Bacteria) ou C-terminal (Archaea e Eukarya). Uma dificuldade encontrada nos estudos estruturais dessas enzimas é o fato de elas serem localizadas inseridas em membrana [67], o que aumenta a complexidade da caracterização cristalográfica necessária para a descrição de estruturas completas.

As primeiras estruturas obtidas foram de subunidades menores do complexo OST de Eukarya, como a Ost4p (2004) [68] e a Ost6L (2009) [69], bem como o primeiro domínio C-terminal de *Pyrococcus furiosus* (2008) (Figura 5) [62]. Em seguida (2009), o domínio Periplasmático (equivalente ao C-terminal em Archaea e Eukarya) de *C. jejuni* foi elucidado (Figura 6) [63]. Posteriormente, os domínios C-terminal das espécies *A. fulgidus* (2012 e 2013 - duas isoformas) [65, 66] (Figura 6 e *Pyrococcus horikoshii* (2013) [70] também foram descritos. Ainda, em 2012 o primeiro domínio C-terminal de Eucariotos foi obtido advindo de *Saccharomyces cerevisiae* (Figura 6) [71]. Mais tarde, uma última estrutura incompleta foi caracterizada, o domínio C-terminal de *Pyrococcus abyssi* (2014).

1.3.3 A primeira estrutura completa: PglB

Em meados de 2011, a primeira estrutura completa de OST foi obtida [64], a PglB de *C. lari* (ClPglB), com resolução cristalográfica de 3,4 Å, considerada um marco no estudo da via de N-glicosilação (Figura 7). A estrutura dessa OST é composta por 712 resíduos de aminoácidos, divididos em dois domínios estruturais, o TM e o PP. O domínio transmembrana engloba os resíduos 1-432, enquanto o periplasmático, os resíduos 433 ao 712. A porção transmembrana dessa proteína envolve 13 hélices inseridas na bicamada, além de duas alças externas (ELs), chamadas EL1 e EL5, que possuem importantes funções relacionadas à catálise da enzima (ver adiante). A EL5 elucidada nessa estrutura possui uma região faltante de 25 resíduos (resíduos 283-306) que correspondem a uma alça aparentemente flexível em sua porção N-terminal. As hélices denominadas TM1 a TM4 e TM10 a TM13 compreendem a região de interface entre os dois domínios, formando assim a cavidade onde ocorrerá a interação com a

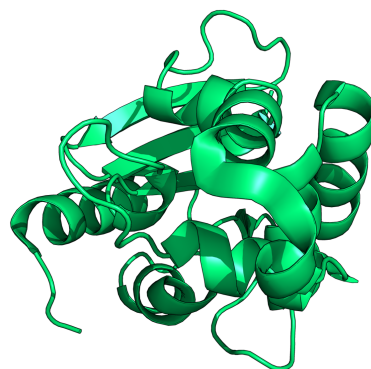


P. furiosus - PfAglB

Figura 5 – Exemplo de organização estrutural encontrada para a AglB de *P. furiosus*. O domínio TM e as unidades estruturais estão coloridas de acordo com o código.



A. fulgidus - AfAglB-S1



A. fulgidus - AfAglB-S2



S. cerevisiae - Stt3

Figura 6 – Exemplos de organizações estruturais semelhantes encontrada para as AglBs S1 e S2 de *A. fulgidus* e para a subunidade catalítica STT3 de *S. cerevisiae*. O domínio TM e as unidades estruturais estão coloridas de acordo com o código.

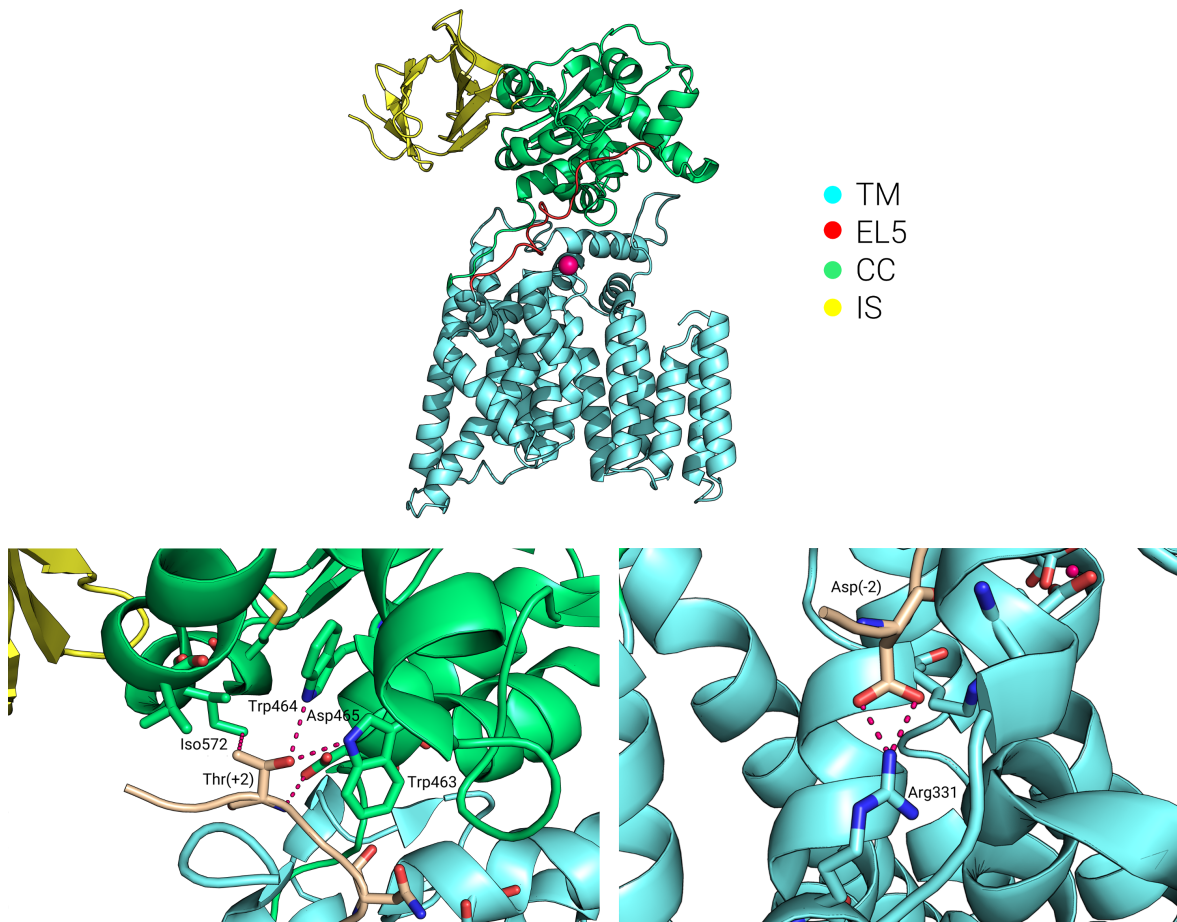


Figura 7 – *Acima*. Estrutura completa da PglB separada por regiões importantes da enzima. *Abaixo*. Resíduos conservados e importantes para o reconhecimento da assinatura da N-glicosilação. O domínio TM e as unidades estruturais estão coloridas de acordo com o código.

proteína aceptora, próximo ao sítio catalítico da PglB. Na face oposta, uma segunda cavidade se apresenta como a região onde supostamente ocorre a interação com o LLO. A conexão entre as duas cavidades é chamada de "estrutura de portinhola" (*porthole structure*), justamente onde a cadeia lateral da asparagina aceptora e o LLO farão contato, estando contidos no sítio ativo da enzima.

Em relação a sua estrutura, o domínio periplasmático possui duas unidades: a CC e a IS. A CC apresenta-se com um enovelamento misto α e β , contendo duas assinaturas principais bem caracterizadas. Essas estão envolvidas na manutenção de uma interação firme da proteína aceptora para que ocorra a catálise. A primeira assinatura foi identificada como WWDYGY (resíduos 463-468), em um primeiro momento [72], e WWDXGX, tardiamente, em decorrência das variações nas tirosinas da assinatura em alguns organismos [47]. O segmento WWD forma o sítio de reconhecimento da posição +2 (S/T) da assinatura da N-glicosilação, interagindo por intermédio de ligações de

hidrogênio entre os resíduos. Além disso, o resíduo Tyr468 foi sugerido como uma peça importante na interação com a N-acetilação do C2 da Bacilosamina após uma primeira modelagem do substrato na cavidade. Outra assinatura importante descrita em bactérias e algumas arqueias é a MI (resíduos 569, 572 e 576) ou, em algumas arqueias e eucariotos, DK. Dados oriundos da estrutura cristalográfica da PglB [64] indicam que essa assinatura explicaria a preferência por uma treonina na posição +2 da assinatura de N-glicosilação, já que a cadeia lateral do resíduo de isoleucina poderia realizar interações hidrofóbicas com a metila presente na cadeia lateral de treoninas e que, por sua vez, está ausente nas serinas. Por fim, outra região já identificada é a DXKK (resíduos 481-484), que parece influenciar as interações com a cadeia oligossacarídica [73]. A unidade IS dessa espécie é composta por nove β -fitas que formam uma estrutura de β -barril distorcido, embora sua função exata permaneça desconhecida.

A elucidação estrutural da PglB completa também possibilitou uma caracterização mais profunda da assinatura estendida de N-glicosilação nas OSTs bacterianas. A Arg331 na CIPglB, um resíduo conservado, situa-se próximo à posição -2 (D/E) da assinatura na estrutura cocristalizada, o suficiente para realizar interações que contribuem na estabilização do substrato proteico dentro da cavidade enzimática. Contudo, estudos posteriores sugeriram [74] que a principal interação nessa região é feita pela Arg147 (CIPglB), sendo que a Arg331 serviria apenas como suporte adicional. Em Bacteria ainda não é claro se o processo ocorre pós-traducionalmente, com uma alça flexível carregando a assinatura, ou co-traducionalmente, tendo a PglB acoplada a unidades translocadoras de proteínas pré-enovelamento. A própria noção da essencialidade da assinatura estendida vem sendo desafiada por descobertas recentes [75, 76], verificando-se que mutações pontuais podem alterar drasticamente a preferência da enzima por diferentes assinaturas, além de observar-se que PglBs de diferentes espécies podem ter preferências por outros resíduos na posição -2.

1.3.4 A segunda estrutura completa: AglB

Em 2013 a primeira estrutura completa de OSTs do domínio Archaea foi obtida: a AglB-L (L = *large*) de *A. fulgidus* (AfAglB-L) [77]. Ela teve sua estrutura caracterizada em dois estados levemente diferentes (Figura 8), um estado nativo (apo) e um estado com um íon sulfato ligado, emulando a presença do fosfato do LLO, com resoluções de 3,4 e 2,5 Å, respectivamente. Outra diferença que chama a atenção é a presença da EL5 completa, previamente não vista na PglB. Posteriormente, em 2017 [78], um estado similar ao da PglB foi determinado (ID 5GMY) na presença de um peptídeo acceptor covalentemente ligado por uma ponte dissulfeto na AfAglB. Apenas duas notáveis diferenças foram

observadas em relação à AfAglB-L previamente caracterizada: a ausência da região N-terminal da EL5 (assim como visto na ClPgLB) e a aproximação do resíduo Glu360 do metal divalente, indicando a importância da interação do substrato acceptor na formação do sítio catalítico completo. A enzima completa possui 868 aminoácidos, divididos nos domínios TM (1-499) e globular C-terminal (500-868). A conexão entre os domínios é realizada por um segmento de dobradiça (resíduos 494-499) e por interações não-covalentes realizadas pela EL1. A estrutura da região TM da AfAglB-L demonstra bastante similaridade estrutural com a ClPgLB, a despeito da sua baixa identidade de sequência (menos de 20%), com organização parecida e mesmo número de hélices (13). A estrutura da EL5 (resíduos 336-373) completa foi detectada em um dos estados, o estado apo, enquanto no estado ligado ao íon, o mesmo grau de desordem visto na PgLB foi identificado, o que poderia indicar que, em estados ligados a diferentes substratos, a porção N-terminal das EL5 de OSTs mostra maior ou menor nível de flexibilidade.

No domínio C-terminal, a AfAglB-L apresenta três unidades estruturais: CC, IS e P1. A unidade CC é similar àquela encontrada na ClPgLB, possuindo os mesmos dois motivos principais de reconhecimento do peptídeo acceptor (WWDYGY e DK/MI). Enquanto isso, a unidade IS difere por ser menor e sem o formato de β -barril encontrado na ClPgLB e em outras estruturas parciais (PfAglB-L e PhAglB-L). Nesse caso, três α -hélices pertencentes à unidade CC parecem substituir o restante da estrutura equivalente ao barril na OST bacteriana. Finalmente, a unidade P1, ainda sem função conhecida, é constituída quase que exclusivamente por β -folhas e, especialmente, tem uma posição similar a estruturas prévias.

1.3.5 O sítio catalítico e o mecanismo de ação das OSTs

A primeira estrutura de OSTs obtida, cristalizada na presença de um peptídeo acceptor portando a assinatura estendida da N-glicosilação de bactérias, permitiu a inferência de um mecanismo catalítico para a enzima (Figura 9). O metal divalente (M^{2+}) presente no sítio catalítico possui um papel fundamental, pois sua ausência abole a atividade enzimática [74]. Na estrutura obtida por Lizak e colaboradores, esse metal era um íon magnésio (Mg^{2+}), abundante na solução cristalina. Todavia, acredita-se que em *C. lari* esse íon seja Mn^{2+} . A natureza do íon no sítio catalítico é inespecífica, pois diversos outros metais são ativos tanto na AgLB [66] como na PgLB [61], sendo essa uma propriedade vista em outras GTases [79]. Na estrutura cristalográfica da PgLB, uma rede de interações foi encontrada no sítio catalítico, com destaque para alguns resíduos específicos: Asp56, Asp154 e Glu319, que coordenam diretamente o íon metálico. Os resíduos Asp56 (localizado na EL1) e Glu319 (localizado na EL5) posicionam-se próximos

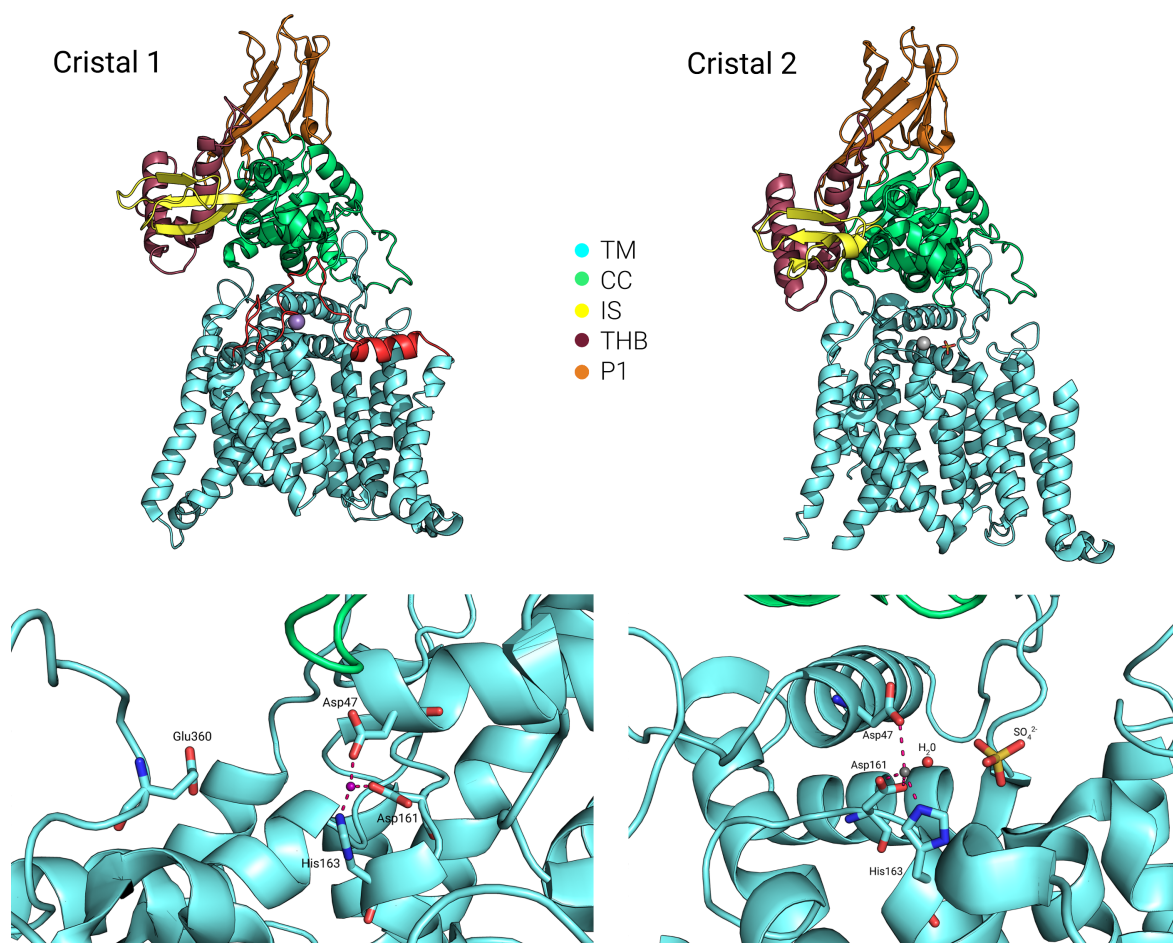


Figura 8 – *Acima*. Estruturas completas da AgIB nos dois estados conformacionais determinados. *Abaixo*. Coordenações distintas do metal divalente encontradas em cada estado no sítio catalítico. O domínio TM e as unidades estruturais estão coloridas de acordo com o código.

a Asn do peptídeo aceptor da glicosilação. O Asp154 (localizado na EL2), importante na coordenação do metal, faz parte de um motivo bastante conservado em GTases, o DXD, juntamente com o Asp156 (sem função definida) [79, 80]. Todos esses resíduos, quando mutados para alaninas, influenciam tanto parcial como totalmente a atividade enzimática da PglB [64, 66]. A hipótese gerada a partir da observação dessas interações e de ensaios bioquímicos [61] (Figura 9) envolve diretamente os resíduos Asp56 e Glu319: os carboxilatos presentes nas cadeias laterais desses resíduos, ao passo que coordenam o metal com um dos átomos de oxigênio, podem formar ligações de hidrogênio com a amida da cadeia lateral da Asn aceptor. A formação dessas ligações induziria a rotação da ligação N-C da carboxamida, o que evitaria a conjugação dos elétrons do nitrogênio com o grupo carbonila, rompendo a ressonância dessa região. Por conseguinte, a hibridização do nitrogênio se tornaria sp^3 e ocorreria uma modificação na posição do par de elétrons, aumentando a eletronegatividade desse átomo e potencializando

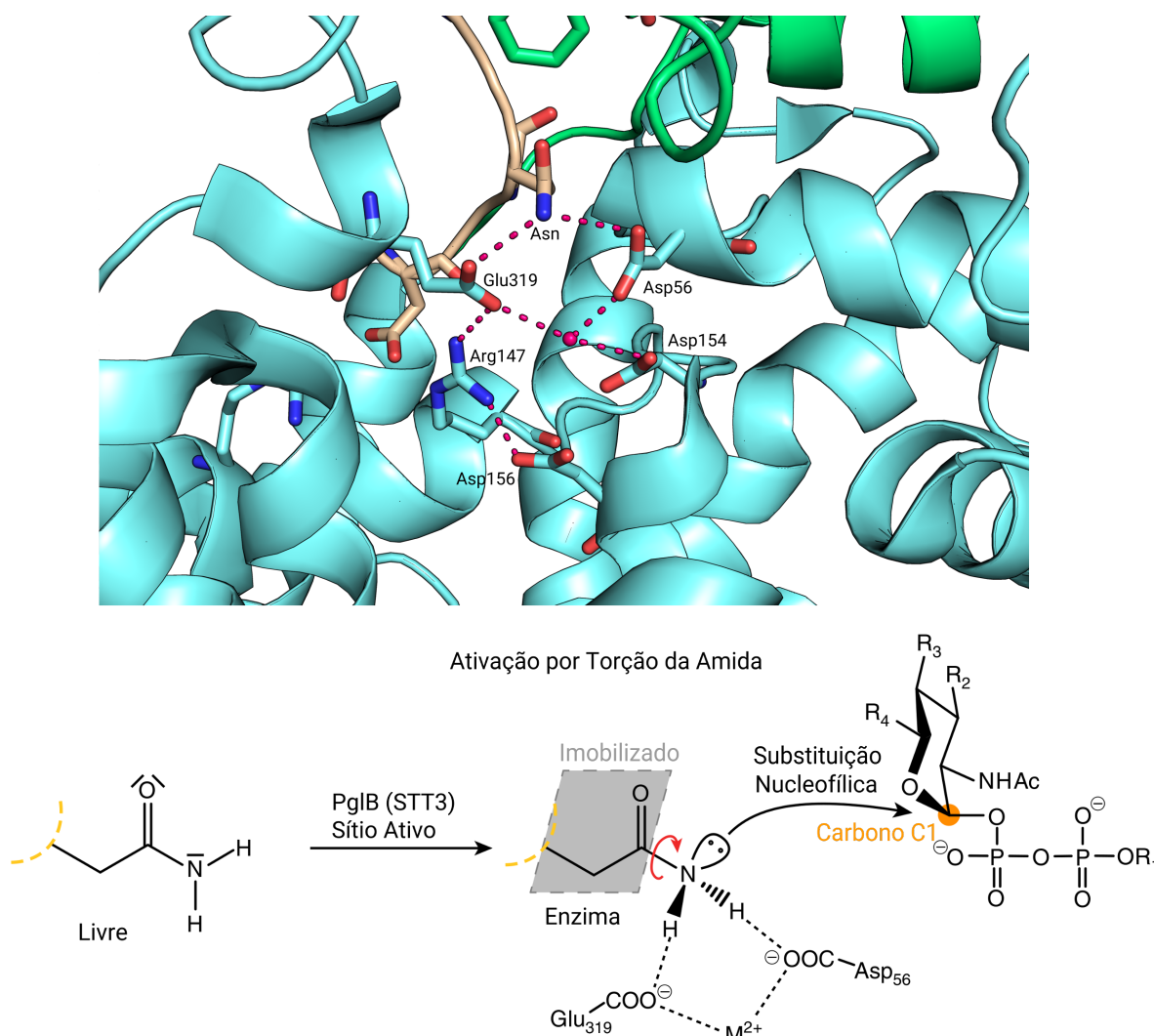


Figura 9 – Rede de interações que regem a coordenação do íon metálico e da cadeia lateral da Asn e proposta de mecanismo de ação baseada na estrutura da PglB. Adaptado de Lizak e colaboradores [61]

sua nucleofilicidade. Esse conjunto de alterações permitiria que o nitrogênio da amida realizasse um ataque nucleofílico no carbono da posição 1 da bacilosamina. O metal possuiria, dessa maneira, uma função dupla - coordenar os resíduos do sítio catalítico e interagir com o pirofosfato do glicolípido para sua estabilização.

Outros elementos participam de forma mais indireta no processo catalítico. Por exemplo, Arg147 e Arg331 interagem com Asp (-2) da assinatura do substrato acceptor. Arg375, altamente conservada, é predita como importante na interação com o pirofosfato [64], auxiliando na manutenção do LLO próximo ao sítio ativo. A alça EL5 funcionaria como duas partes independentes: a porção N-terminal parece ser altamente flexível na presença do acceptor proteico, ou ordenada na forma livre da enzima; enquanto isso, a sua porção C-terminal na presença do acceptor é rígida e se acopla à estrutura da

enzima com Glu319, para formar o sítio ativo, desordenando-se na sua forma livre [81]. O resíduo Tyr293 do motivo Tyr-*plug*, pertencente à porção N-terminal da EL5, parece ser importante para a catálise, talvez interagindo com o LLO e auxiliando a reação, porém não é essencial para a ligação do substrato proteico na enzima [81]. Entretanto, seu papel exato ainda é desconhecido.

Observou-se na AglB há algumas modificações nesse processo que ainda não são compreendidas. O metal encontrado na estrutura cristalografada foi Zn^{2+} , porém foi sugerido que aquele que ocorre naturalmente seria o Mg^{2+} [79]. De qualquer maneira, a enzima é cerca de 3x mais ativa na presença de Zn^{2+} do que em outros metais, indicando que esse metal pode ser um bom substituto [77]. Os resíduos que no cristal parecem interagir com o metal são Asp47 (equivalente a Asp56 na ClPglB), Asp161 (equivalente a Asp154 na ClPglB) e His163 (equivalente a Asp156 na ClPglB). O resíduo Glu360 não coordena o metal como seu equivalente (Glu319) na ClPglB, reforçando a ideia de que o acoplamento da porção C-terminal de EL5 responde à presença (ClPglB) ou à ausência (AfAglB) do substrato proteico [81]. No lugar dessa interação, uma molécula de água foi encontrada assumindo a função de coordenação. Na AglB, o íon sulfato encontrado posicionado próximo ao sítio ativo (Figura 8 - cristal 2) foi hipotetizado como emulando o fosfato do dolicol carreador, o que poderia indicar que esse era um estado pós-catalítico [77]. Ao estudar algumas mutações nos resíduos próximos a esse íon (His81, His162, Trp215 e Arg426), foi verificado que nas histidinas a inversão das cargas aboliu a atividade, enquanto a mudança de uma arginina para alanina teve o mesmo efeito [77]. Nesse estado, com a presença do íon sulfato, a porção N-terminal da EL5 está desordenada e não tem sua estrutura elucidada, sugerindo que a presença do lipídeo-carreador também influencia na estrutura da alça [77].

1.4 Aplicações biotecnológicas da via de N-glicosilação e a glicoenharia de proteínas

A elucidação dos sistemas de N-glicosilação em Bacteria e em Archaea permitiu a operacionalização de aplicações biotecnológicas envolvendo a manipulação dessas vias, tanto intra como interespecies (Figura 10, alimentando o desenvolvimento da glicoenharia de proteínas [82, 83]. A importância de investigações envolvendo esse campo pode ser ilustrada pela grande quantidade de estudos empregando proteínas glicosiladas que já estão aprovadas ou que estão em fase de desenvolvimento clínico ou pré-clínico nas agências regulatórias Européia e Americana para uso em humanos [84]. Em muitas dessas proteínas glicosiladas a glicana é essencial para a função ou influencia

diretamente na farmacocinética desses biofármacos. Outra aplicação promissora diz respeito ao estudo de métodos alternativos na produção de vacinas, pois a tecnologia de acoplamento de glicanas a proteínas (*Protein glycan coupling technology* - PGCT) demonstra diversas vantagens sobre o método empregado atualmente (conjugação química) [85]. Baseados nessa tecnologia, alguns artigos na literatura vêm descrevendo abordagens envolvendo o uso de enzimas pertencentes à maquinaria de glicosilação bacteriana para, por exemplo, realizar a adição de diferentes tipos de estruturas sacarídicas, como os O-antígenos de espécies bacterianas, para auxiliar no desenvolvimento de vacinas [86–89]. Uma forma mais recente da aplicabilidade em vacinas envolve o uso da via de N-glicosilação de *C. jejuni* transfectada para linhagens não-patogênicas de *Escherichia coli* [90]. Nesse caso, a célula bacteriana produz vesículas de sua membrana externa que apresentam a glicana. Essas vesículas foram utilizadas na vacinação de galinhas, que demonstraram uma redução de 10^4 na colonização bacteriana. Em termos de pesquisas, uma boa parte dos esforços envolvem estudos que visam à produção de diferentes cadeias oligossacarídicas existentes em eucariotos. Já foram obtidos, com sucesso, o núcleo inicial trimanosídico de mamíferos (juntamente com a quitobiase), a sua transferência pela PglB para outras proteínas [91], bem como o antígeno Lewis X [92,93]. Paralelamente, outras biomoléculas glicosiladas também vêm sendo estudadas como, por exemplo, o lipídio A, componente do lipopolissacarídeo (LPS) presente na membrana bacteriana, que recebeu modificações para a formação de LPS quiméricos. Esses, por sua vez, possuem estruturas como a de gangliosídeos e de antígenos Lewis (X e Y) [94–96].

Um dos principais fatores que torna as OSTs plataformas promissoras para estudos de glicoengenharia é a promiscuidade dessas enzimas, especialmente a PglB de *C. jejuni*, somada ao fato de possuir apenas uma subunidade e de poder ser transferida para um organismo que não possua um sistema nativo de N-glicosilação, como *E. coli* [30]. Reforçando essa informação, estudos prévios [97,98] verificaram que um dos únicos requisitos para uma glicana qualquer ser transferida pela PglB é a presença de um grupo acetamida na posição C-2 do monossacarídeo que será ligado na asparagina acceptora. Pesquisas recentes estão sendo direcionadas para modificações que alterem tanto a plasticidade do sítio ativo da PglB [88] como as assinaturas de N-glicosilação, buscando menor especificidade e maior eficiência no processo.

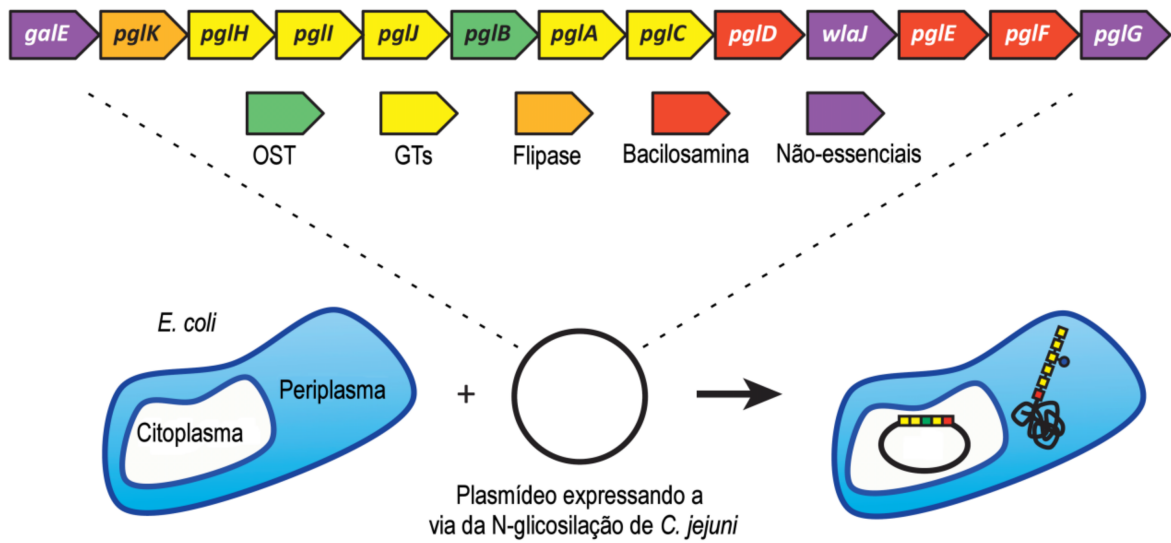
Como toda área em desenvolvimento, a glicoengenharia possui desafios que limitam a propagação de suas aplicações [82]. É necessário maior detecção e caracterização de novas estruturas oligossacarídicas, especialmente em bactérias e archaeas, onde ocorre a maior variabilidade, bem como produção e conjugação mais eficientes de glicanas eucarióticas em bactérias, visando à expansão das ferramentas disponíveis. Também é

importante encontrar um melhor entendimento das vias de N-glicosilação, visto que elas possuem variações interespecíficas, propiciando assim uma manipulação melhor 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 das proteínas-alvo, possibilitará avanços no desenvolvimento e no desenho específico de glicanas com diferentes aplicações.

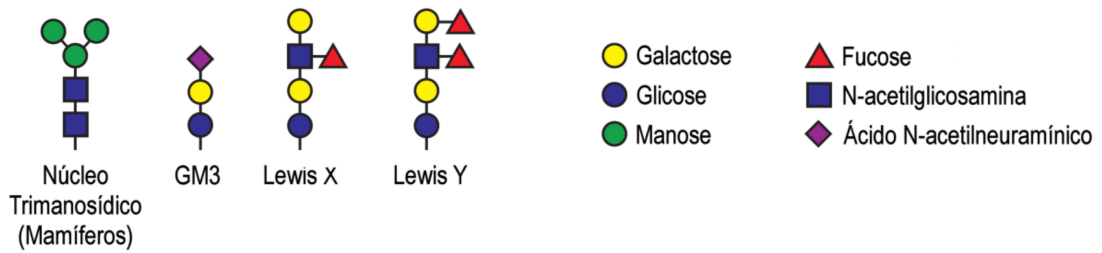
1.5 Elucidação da estrutura e da conformação de biomoléculas

A área de estudo envolvendo a determinação de estruturas de biomoléculas surgiu em meados do século XX e desde então vem agregando contribuições expressivas para o avanço científico [99]. Ao realizar a caracterização estrutural de uma molécula em nível atômico, é possível acessar suas interações moleculares, o que, potencialmente, trará informações que auxiliam na inferência da sua função [100]. A compreensão das bases estruturais que definem o papel desempenhado por diferentes moléculas facilita a descrição de fenômenos biológicos em níveis variados, como a inibição causada por pequeno ligante em um receptor, o reconhecimento intercelular mediado por glicoproteínas, o transporte através de membranas, o funcionamento de ribossomos e a formação de capsídios virais [100]. Nesse contexto, diferentes técnicas vêm sendo desenvolvidas visando a aplicações na elucidação estrutural de entidades biológicas em resoluções distintas. Algumas técnicas são informativas em termos de características globais, como, por exemplo, o espalhamento de raios-X a baixos ângulos, que proporciona a obtenção de dados relativos à forma, ao tamanho e à distribuição de partículas [101]. Já a técnica de difração de raios-X é capaz de descrever o enovelamento proteico e o conteúdo de estrutura secundária [102]. Entretanto, essas técnicas não são capazes de descrever com resolução atômica as biomoléculas que são estudadas.

As principais técnicas de escolha para a determinação de estruturas tridimensionais de biomoléculas com resolução atômica, atualmente, são a ressonância magnética nuclear (RMN) e a cristalografia de raios-X (Figura 11). Entre as duas, a cristalografia é a técnica dominante, sendo que das mais de 130.000 entradas no banco de dados do *Protein Data Bank* (PDB), cerca de 90% das estruturas foram resolvidas utilizando esse método (acessado em junho de 2017). Essa abordagem baseia-se na propriedade de agregação das moléculas, organizadas em arranjos periódicos que correspondem às estruturas de cristais [103, 104]. A ocorrência da cristalização é influenciada por condições variadas, como o pH, a concentração salina usada e a temperatura empregada, fatores que podem determinar o sucesso do processo [105]. Apesar de ser o método mais frequentemente empregado e ter sido amplamente refinada desde o seu início, ainda



Exemplos de glicanas já produzidas em bactérias:



Genes produzindo glicanas de superfície de bactérias patogênicas

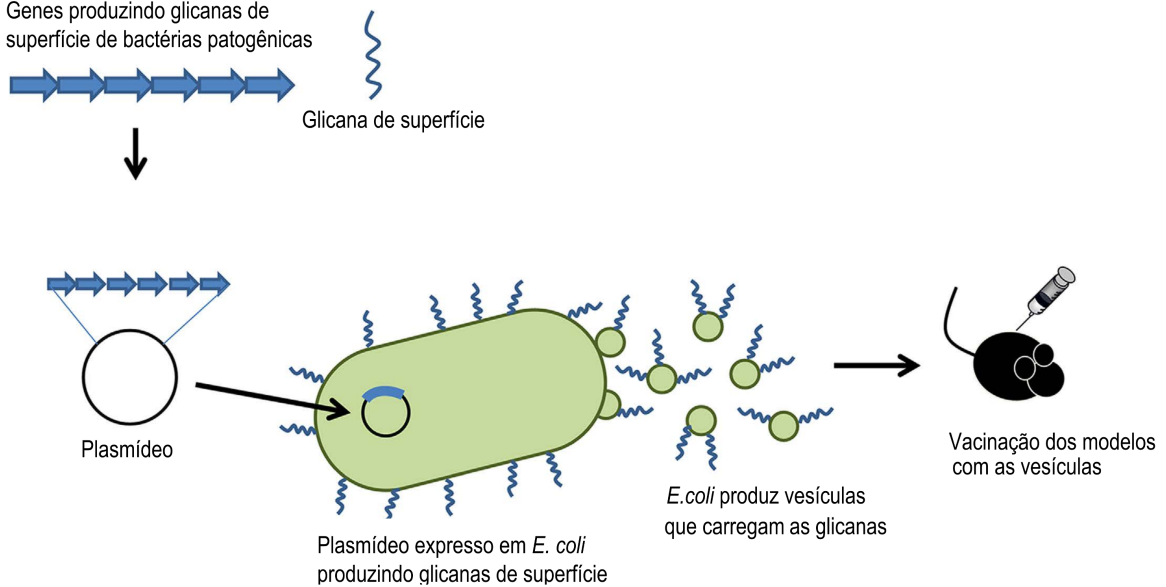


Figura 10 – *Acima*. Representação esquemática da glicoengenharia empregada na expressão em *E. coli* Adaptado de Baker, 2013 [82]. *Abaixo*. Exemplo de aplicação no desenvolvimento de vacinas pela produção de vesículas da membrana externa apresentando a glicana ao hospedeiro. Adaptado de Price *et al.*, 2016 [90].

apresenta algumas limitações [?]: i) a formação dos cristais contendo a molécula de interesse, devido à multiplicidade de fatores que participam no processo; ii) a obtenção de proteínas de membranas, em decorrência do ambiente em que está inserida; iii) a indução de estabilidade estrutural pelo ambiente cristalino, que pode acarretar estruturas com erros em suas definições; iv) impossibilidade de se obter um conjunto de conformações, sendo a estrutura resolvida um estado médio.

O método de RMN, em contrapartida, oferece uma caracterização que permite observar a flexibilidade de regiões da estrutura e os diferentes estados conformacionais das proteínas [106]. Nesse método é possível identificar a presença de núcleos de átomos que estão espacialmente próximos, ou seja, na vizinhança um do outro, gerando mapas de restrição espacial que serão utilizados para a construção dos modelos tridimensionais. O efeito Overhauser nuclear (NOE, *nuclear overhauser effect*) - que ocorre quando átomos próximos estão em estados diferentes de magnetização - é empregado em espectros de resolução bidimensional [107] com o intuito de detectar a proximidade entre átomos não-covalentemente ligados (menor que 5 Å) [108]. Essa técnica, porém, apresenta limitação quanto ao tamanho das moléculas que podem ser analisadas, com os maiores modelos chegando a 1 mega Dalton de peso molecular atualmente [109, 110], em decorrência da sobreposição de picos nos espectros utilizados na construção dos modelos estruturais, impedindo a identificação correta da posição atômica. Assim como a cristalografia, a RMN também enfrenta restrições na capacidade de elucidar estruturas de proteínas de membrana, tendo em vista que ocorrem perdas de estrutura e função dessas moléculas ao serem retiradas do ambiente lipídico e serem purificadas em detergentes [106].

Uma mudança de paradigma que pode estar acontecendo diz respeito ao uso da técnica de microscopia crioeletrônica, cujo emprego na determinação estrutural vem aumentando consideravelmente (em 2015 = 216, em 2016 = 410) [111]. A técnica funciona bem para estruturas de tamanho superior a 200 kDa, como complexos macromoleculares e para proteínas transmembranas grandes. Todavia, tal abordagem ainda enfrenta um problema de resolução, que dificilmente atinge valores próximos aos da cristalografia (Figura 11, apesar de algumas estruturas já atingirem resoluções menores que 3 Å [112, 113]). Outra questão diz respeito à flexibilidade da molécula estudada: ela pode tanto ser descrita corretamente, conferindo um aspecto positivo ao método, como pode impedir que a caracterização de determinadas regiões seja realizada [111].

Tratando especificamente do tema dessa tese, os glicoconjugados são um caso à parte quando se fala em estudos estruturais. Os dados relativos a essas biomoléculas são usualmente escassos, particularmente por causa das dificuldades encontradas na caracterização de suas estruturas. O principal fator diz respeito à alta flexibilidade

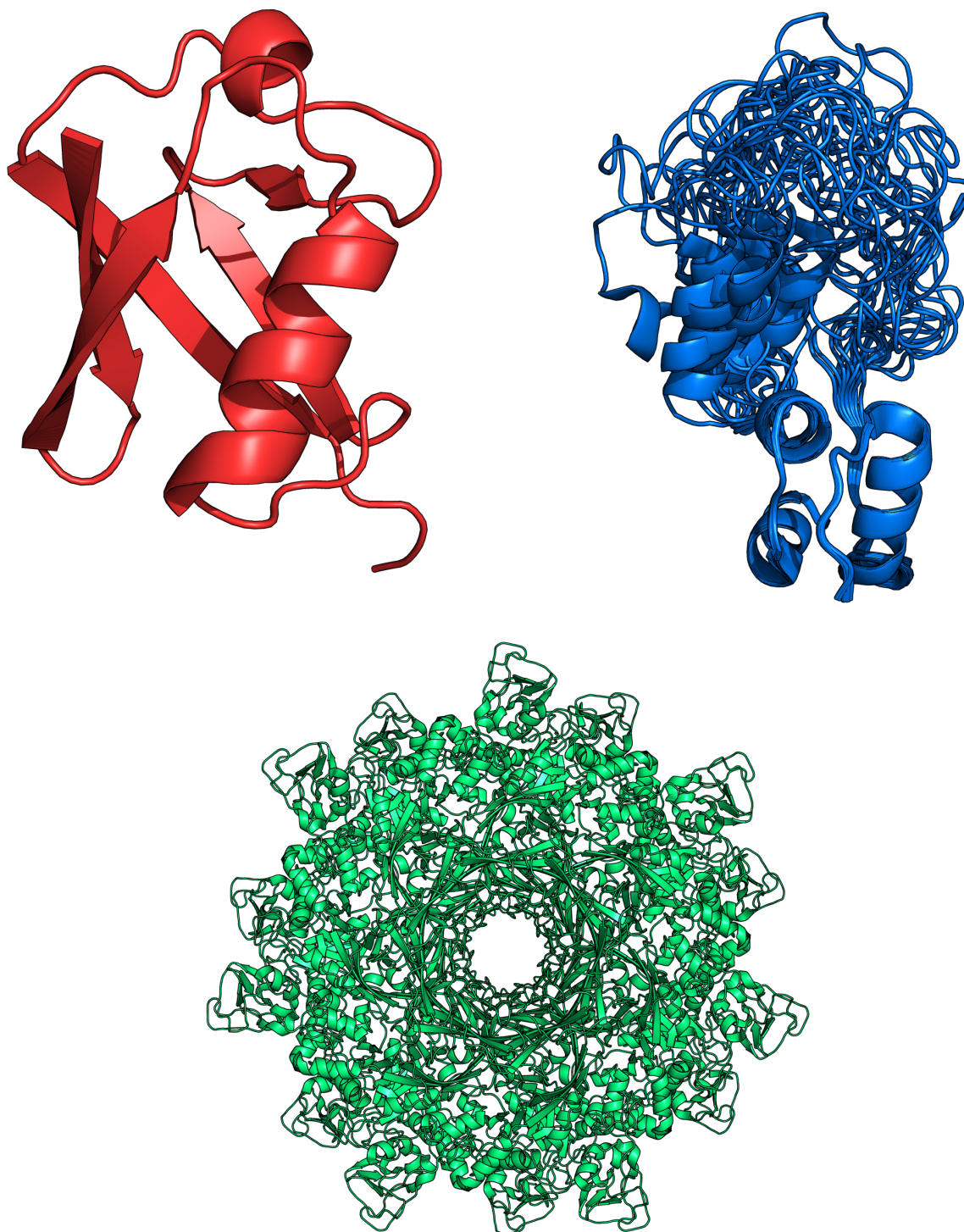


Figura 11 – Estruturas tridimensionais obtidas pelos diferentes métodos disponíveis atualmente. *Vermelho*. Cristalografia de Raio-X - Ubiquitina (ID 1UBQ) [114]. *Azul*. Ressonância Magnética Nuclear - Proinsulina (ID 2KQP) [115]. *Verde*. Microscopia Crioelétrica - Aerolisina (ID 5JZH) [116].

conferida pela presença dos componentes glicídicos. Na cristalografia, essa flexibilidade atrapalha a formação dos cristais [117–119]. Adicionalmente, a hidroflicidade que gera um aumento na quantidade de moléculas de água coordenando suas estruturas, bem como a ausência de interações hidrofóbicas e dipolares entre resíduos sacarídicos próximos, são também fatores impeditivos [120]. Por fim, nos casos das glicoproteínas, seus mapas de densidade eletrônica são prejudicados por essa flexibilidade exacerbada, interferindo na correta determinação estrutural dessas moléculas [121]. Já no método de RMN, a mobilidade dessas cadeias sacarídicas produz dificuldades na identificação de múltiplos sinais de NOE. A baixa quantidade de sinais obtidos, seja entre ligações glicosídicas, seja entre a própria proteína com o carboidrato, pode gerar a obtenção de estados conformacionais médios [119,120], que não representam adequadamente os estados existentes em solução. Quanto ao método de microscopia crioeletrônica, esse pode ser um bom aliado em estudos de macrocomplexos glicosilados. Porém algumas questões já mencionadas ainda permanecem, como a dificuldade em lidar com moléculas menores e flexíveis, além de eventuais problemas na obtenção de resolução apropriada. Acrescenta-se ainda o fato de que é estimado que cerca de um terço das estruturas de carboidratos depositadas no PDB apresentam algum tipo de inconsistência [122,123].

Nesse contexto, os métodos computacionais de avaliação da estrutura e da dinâmica de biomoléculas funcionam como alternativas aos métodos tradicionais [124–126], especialmente em glicoconjugados [127–133]. Dentre essas abordagens computacionais, tem-se a dinâmica molecular (DM), técnica que foi introduzida na década de 70 [134,135] e que vem sendo utilizada atualmente para problemas diversos [136–139]. Essa abordagem proporciona muitas vezes subsídios para novos experimentos de maneira única, beneficiando-se da sua resolução atômica e da forma singular com que possibilita a observação de propriedades tempo-dependente. A DM irá avaliar o comportamento físico-químico de um determinado sistema de átomos, baseando-se em forças que podem induzir a movimentação desse sistema. Para tanto, o método realiza sucessivas integrações da equação de movimento de Newton aplicada a todos os átomos do sistema, empregando um tempo de integração denominado dt [140]. Isso é realizado para todas as posições e interações dos átomos durante uma quantidade de tempo, sendo atualizado constantemente e gerando o que é chamado de trajetória. Essa trajetória produzida contém um conjunto de conformações (*ensemble*), as quais descrevem o perfil conformacional encontrado para as moléculas estudadas [100]. A DM está atrelada ao conceito da mecânica molecular (MM), o que implica na consideração apenas dos núcleos atômicos nos cálculos, excluindo-se a descrição direta de propriedades eletrônicas, como a formação e a quebra de ligações atômicas.

Mais especificamente, a DM é regida pela seguinte equação: $F_i = m_i \cdot a_i$. O termo

a_i pode ser descrito como $a_i = d^2r_i(t)/dt^2$, onde $r_i(t)$ corresponde à posição do átomo i e dt ao tempo de integração discretizado [140]. O termo relativo à força do sistema também pode ser descrito de outra forma: $F_i = -dV/dr_i$, onde V diz respeito à energia potencial do sistema calculada para cada átomo i na posição r_i . Esse ponto é importante para a determinação da intensidade e da direção da F_i , uma vez que esse termo é colocado em função das coordenadas cartesianas dos átomos e da energia potencial do sistema. Dessa forma, a equação é integrada considerando a variação temporal (t); as forças agindo sobre o átomo i são calculadas pelo conjunto de interações descritas para esse átomo (campo de força); essas forças geram uma variação no espaço cartesiano (dr), em função do tempo de integração (usualmente 1 ou 2 fs). Ao resolver essa equação para cada átomo do sistema é possível obter a mencionada trajetória, onde o comportamento do sistema durante um tempo de simulação será avaliado [140].

Para calcular a energia potencial em relação aos átomos que compõem o sistema estudado são necessários os chamados termos ligados e termos não-ligados. A esse conjunto de parâmetros e de funções é dado o nome de campo de força (Figura 12). A forma funcional básica desses campos de força é dada pelo somatório desses termos. Alguns campos de força mais sofisticados podem apresentar termos adicionais, entretanto, os termos aqui apresentados estão invariavelmente presentes [140]. Os termos ligados avaliados são: termos de ligação covalente entre dois átomos ($V_{\text{ligação}}$), descritos por um potencial harmônico (k_b) que penaliza energeticamente variações de b em relação ao valor de referência b_0 ; termos de descrição de ângulos envolvendo três átomos ($V_{\text{ângulo}}$), que de forma semelhante emprega um potencial harmônico (k_θ) para evitar variações excessivas, onde θ é o ângulo calculado e θ_0 é o ângulo de referência; termos torsionais relativos à rotação de ângulos diedrais formados entre quatro átomos (V_{diedrais}). O diedro próprio é descrito por uma função cosseno com n sendo o valor de periodicidade ou multiplicidade (indicando o número de mínimos de energia), ϕ é o valor do diedro, k_ϕ determina a barreira energética para que ocorra mudança de mínimo de energia e δ indica o máximo de energia no perfil rotacional do diedro. O diedro impróprio varia para cada campo de força, podendo tanto seguir o padrão periódico encontrado para diedros próprios (AMBER e OPLS), como o padrão de potenciais restritivos encontrado para ligações e ângulos (CHARMM e GROMOS). Os termos não-ligados compõem a soma de dois outros termos: os termos coulômbicos relativos a componentes eletrostáticos dos átomos (V_{eletro}), calculados como a soma de interações entre pares de cargas atômicas, utilizando a Lei de Coulomb; os termos de van der Waals (V_{LJ}), calculados aplicando um potencial de Lennard-Jones 12-6 entre pares de átomos neutros. Apesar dessa forma básica, cada campo de força tem suas peculiaridades para lidar com a resolução dessas equações. Atualmente, os campos de força mais comumente utilizados são o Amber [141],

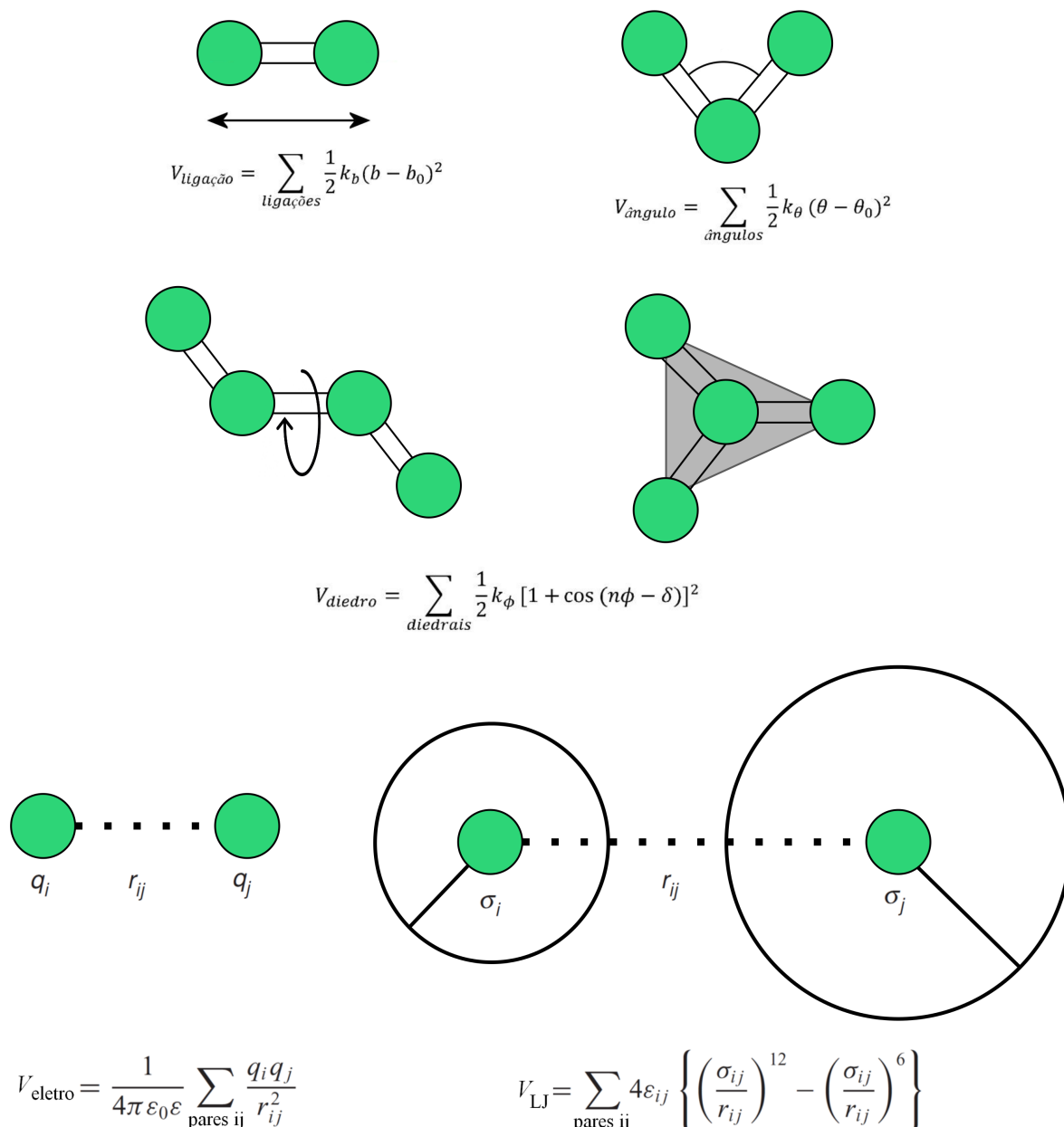


Figura 12 – Esquema representativo dos parâmetros que descrevem a forma funcional de um campo de força.

CHARMM [142], GROMOS [143] e OPLS-AA [144].

1.5.1 Validação das simulações por DM

As simulações em DM apresentam, de modo geral, duas limitações principais: a precisão dos campos de força na descrição das biomoléculas de estudo e a amostragem suficiente das simulações. A escolha do campo de força pode influenciar na correta caracterização dos sistemas estudados, entretanto não há um consenso sobre um campo de força que seja superior aos outros em todos os quesitos. Já a amostragem está

ligada diretamente à questão da convergência e da reproducibilidade, onde, idealmente, diferentes experimentos computacionais deveriam observar comportamentos similares ou convergentes. Todavia, a convergência não é um fator facilmente evidenciável [145], sendo ainda um problema [146–148], especialmente para proteínas maiores [149], mesmo quando as simulações são executadas por períodos mais longos. Dessa forma, alguns grupos propõem o conceito de convergência relativa [150], enquanto outros [148] consideram que uma simulação foi validada e amostrada suficientemente quando consegue responder questões científicas específicas. Uma forma de diminuir os problemas de pouca amostragem é o emprego de replicatas das simulações [151], utilizando velocidades iniciais diferentes para garantir que o sistema simulado percorra um caminho distinto dos outros. Outra maneira envolve o uso de métodos de amostragem ampliada, como a metadinâmica [152] e a *replica-exchange* [153]. Nesses métodos, o sistema é forçado, por metodologias distintas, a buscar novos estados conformacionais, não permanecendo preso apenas a estados energeticamente favoráveis.

A validação dos dados advindos dessas DMs com variadas abordagens pode ser feita pela comparação com dados experimentais disponíveis na literatura. Se fenômenos biológicos forem respondidos, explicados ou observados, isso pode indicar que a simulação contém resultados apropriados. Outra forma é a verificação de propriedades físico-químicas experimentais das biomoléculas (ex: entalpia de vaporização, capacidade térmica) sendo estudadas, servindo como uma validação do experimento computacional.

1.5.2 Aplicação de abordagens computacionais para o estudo de OSTs

No contexto de estudos estruturais, a DM emerge como uma forma mais conveniente de se obter uma noção das propriedades dinâmicas de biomoléculas, permitindo a obtenção de perfis conformacionais [100]. Nesse sentido, quando reconstruídos de maneira apropriada, esses perfis podem fornecer informações a respeito de eventos conformacionais completos, como a transição entre estados biologicamente relevantes ou eventos de enovelamento [154, 155]. Esses conjuntos de conformações também possuem a vantagem de propiciar o cálculo de médias de propriedades do conjunto, ao invés de utilizar valores únicos gerados por estados conformacionais médios [156, 157]. Dessa forma, a utilização dessa técnica para o estudo de OSTs pode auxiliar na compreensão dos mecanismos que envolvem esse sistema.

Até o presente momento, apenas três trabalhos na literatura empregaram essa técnica no estudo desses sistemas. O primeiro deles, publicado em 2014 por Kern e colaboradores [158], realizou a descrição conformacional de dois LLOs, o glicolípido bacteriano portando a glicana de *C. jejuni* e o glicolípido eucariótico portando uma

glicana conservada de eucariotos. Esse trabalho trouxe informações importantes acerca do comportamento estrutural do LLO em membrana e propôs uma conformação de interação com a PglB para o LLO bacteriano.

O segundo trabalho foi realizado pelo Grupo de Bioinformática Estrutural e diz respeito à simulações por DM da CIPglB inserido em um modelo de membrana biológica na presença de solvente explícito e de substratos distintos. Esse trabalho foi iniciado durante o meu Mestrado [159], onde as primeiras simulações do sistema foram realizadas, dando enfoque às alterações globais da estrutura causadas pelos diferentes substratos, bem como a uma primeira caracterização da dinâmica do sítio ativo. Posteriormente, dada a complexidade do tema abordado, esse trabalho continuou sendo desenvolvido no Doutorado, culminando em publicação em periódico [160] que será abordada no Capítulo de Resultados dessa tese.

O terceiro trabalho é bastante recente (30 de Maio de 2017) [161] e compreende a simulação por DM do complexo completo da PglB, ou seja, contendo tanto o peptídeo acceptor e como LLO, assim como a PglB na presença de outros substratos. O artigo de Lee & Im apontou movimentos do domínio TM como os responsáveis pela modulação da EL5 e, por conseguinte, da indução de alterações conformacionais relevantes para a atividade da OST.

A despeito desses trabalhos terem abordado questões relativas à estrutura das OSTs, fazendo inferências diversas sobre seus mecanismos, muitas questões ainda permanecem não resolvidas sobre as bases moleculares dessas enzimas. Alguns desses pontos incluem: a compreensão do mecanismo de ação da enzima; a importância das diferentes unidades estruturais espalhadas pelas diferentes OSTs dos domínios da vida; os requerimentos para o processo catalítico (como a N-acetilação no carbono da posição 2 do monossacarídeo da extremidade redutora do LLO); a contribuição de resíduos específicos da PglB para a atividade. A complexidade de tratar sistemas incluindo uma proteína transmembrana, a própria membrana biológica, o solvente explícito e diferentes substratos adiciona mais alguns desafios aos estudos, trabalhando como um fator limitante para a amplitude de investigações sobre o tema.

2 Justificativa

Enzimas do grupo das OSTs, como a PglB e a AglB, vêm se tornando importantes alvos de investigação para diversas aplicações. Por serem enzimas promíscuas, processando substratos distintos daqueles encontrados naturalmente em seus organismos de origem, a caracterização dos mecanismos envolvendo suas atividades é de grande relevância para a produção de glicoproteínas específicas.

Estudos em nível atômico de sistemas moleculares envolvendo diferentes classes de biomoléculas atuando em conjunto, como proteínas, lipídios e carboidratos, são desafiadores e poucas técnicas experimentais conseguem realizar uma descrição apropriada de seus comportamentos. Nesse sentido, os métodos computacionais representam - muitas vezes - uma alternativa de custo acessível, rápida e com resolução atomística. Ainda, abordagens desse tipo são um dos poucos métodos que permitem o estudo conformacional de carboidratos, já que tanto o RMN como a cristalografia tem dificuldades ao lidar com esse tipo de biomolécula.

Dessa forma, a compreensão de como diferentes OSTs comportam-se, tanto na presença de diferentes substratos como ao sofrerem modificações estruturais, pode contribuir para um melhor entendimento das bases moleculares que regem seus processos catalíticos. Esse conhecimento pode ser extrapolado para enzimas similares e auxiliar em estudos caso a caso, fornecendo subsídios para futuros ensaios experimentais ou para potenciais aplicações biotecnológicas/industriais, como o desenvolvimento de vacinas e a glicoengenharia bacteriana.

3 Objetivos

O objetivo desse trabalho é analisar o comportamento molecular das OSTs PglB e AglB na presença de diferentes condições. A presença ou a ausência tanto do substrato proteico como do substrato lipídico, a presença ou a ausência de suas diferentes unidades estruturais e de seus perfis conformacionais serão avaliadas. Para atingir esses objetivos, o trabalho foi dividido em 3 capítulos, cada um com suas metas:

- I. Descrição da dinâmica do sítio catalítico da PglB de *C. lari*:
 - i. Caracterização da dinâmica de coordenação do Mg^{2+} e de resíduos catalíticos;
 - ii. Analisar a influência conformacional da complexação dos substratos proteico e glicoproteico.
- II. Avaliação do papel funcional das unidades estruturais da AglB de *A. fulgidus*:
 - i. Avaliar a estabilidade das AglBs mutantes ($\Delta P1$ e $\Delta ISP1$);
 - ii. Verificar se há manutenção dos requisitos básicos para catálise em todos sistemas;
 - iii. Identificar as principais interações que regem a organização de unidades estruturais.
- III. Caracterização dos estados pré- e pós-catalíticos da PglB de *C. lari*:
 - i. Parametrizar o LLO e analisar como a sua presença afeta a organização do sítio ativo e dos resíduos relevantes no seu reconhecimento;
 - ii. Obter dois estados catalíticos relevantes que serão empregados em cálculos de mecânica quântica (QM) para realizar a catálise da PglB.

4 Procedimentos metodológicos

*“Remember that all models are wrong;
the practical question is how wrong
do they have to be to not be useful”*

George Box

Buscando a resolução das questões estabelecidas nos objetivos dessa tese, o trabalho foi dividido em três capítulos. O Capítulo I abrange estudos da ClPglB inseridos em membranas biológicas (1-palmitoil-2-oleil-sn-glicero-3-fosfoetanolamina - POPE) avaliando o comportamento estrutural e dinâmico na sua forma apo (nativa), na presença do peptídeo acceptor que foi cocristalizado com sua estrutura e na presença de um glicopeptídeo de *C. lari* e um de *C. jejuni*. O Capítulo II discorre a respeito das unidades estruturais de OSTs, empregando como modelo de estudo a outra estrutura já cristalografada (AfAglB) inserida em POPE, realizando inferências funcionais para essas porções ainda pouco estudadas. Por fim, o Capítulo III retorna a ClPglB inserida em POPE, compreendendo a parametrização do substrato doador - LLO, bem como a formação do complexo completo dessa OST, sendo apresentada uma análise da influência desse substrato no corpo da enzima. A lista completa dos sistemas estudados segue abaixo:

- ClPglB-Nativa-POPE (Capítulo I)
Sistema da PglB nativa utilizado para verificar como a ausência de substratos influencia globalmente a estrutura da enzima, também servindo como um controle;
- ClPglB-Peptídeo-POPE (Capítulo I)
Sistema da PglB na presença de um peptídeo acceptor que porta a assinatura de N-glicosilação. Obtido diretamente da estrutura cristalográfica, funcionou como um primeiro modelo de avaliação do comportamento de elementos catalíticos importantes;
- ClPglB-Glicopeptídeo-POPE (Capítulo I)
Sistemas da PglB na presença de glicopeptídeos, separadamente, de *C. lari* e *C. jejuni* visando analisar as mudanças causadas pela formação do produto da reação da OST;

- AfAglB-L-WT-POPE (Capítulo II)
Sistema da AglB nativa obtida diretamente da estrutura cristalográfica, funcionando como o controle dos experimentos em comparação com as mutações;
- AfAglB-L- Δ P1-POPE (Capítulo II)
Sistema da AglB recebendo a deleção da unidade estrutural P1, com objetivo de avaliar o papel estabilizador dessa unidade;
- AfAglB-L- Δ ISP1-POPE (Capítulo II)
Sistema da AglB recebendo uma dupla deleção das unidades estruturais IS e P1, com a intenção de verificar o papel estabilizador dessas unidades na OST;
- LLO-POPE (Capítulo III)
Sistema envolvendo o Glicolípido de *C. lari* inserido em membrana empregado para a obtenção de um modelo de substrato doador que foi posteriormente complexado na CIPglB;
- CIPglB-Pré-POPE (Capítulo III)
Sistema da PglB em seu estado pré-catálise portando o peptídeo acceptor e o LLO para que fossem estudadas as modificações estruturais induzidas por esses substratos na PglB, juntamente com a caracterização de conformações pré-catalíticas;
- CIPglB-Pós-POPE (Capítulo III)
Sistema da PglB em seu estado pós-catálise portando o glicopeptídeo e a undecaprenila para realizar a caracterização de conformações pós-catalíticas;
- CIPglB+P1-POPE
Sistema da PglB quimérica contendo uma unidade estrutural P1 de *A. fulgidus* ligada no C-terminal de sua estrutura, em uma tentativa de conferir termoestabilidade para a enzima bacteriana.

4.1 Construção dos sistemas

A PglB possui uma estrutura obtida pelo método de cristalografia por raios-X, depositada no banco de dados do PDB, sob código 3RCE [64]. Como comentado na Introdução, as estruturas cocrystalografadas na presença do peptídeo acceptor possuem a região denominada EL5 onde faltam aminoácidos em sua porção N-terminal em relação a sua sequência original. Na PglB, esses resíduos correspondem ao de número 283 até o de número 307 (25 AA). Além disso, há um pequeno trecho de 3 aminoácidos (605-607) na unidade IS. Essas regiões foram preenchidas a partir da submissão dessa estrutura

ao servidor FALC-Loop [162], que utiliza uma metodologia [163] 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 faltantes e é considerada importante para o mecanismo da catálise, a seleção do melhor modelo foi feita buscando essas características. Assim, dentre as 100 estruturas de EL5 resolvidas pelo servidor, o modelo selecionado foi aquele que possuiu melhor score na avaliação da estereoquímica (analisado com o servidor PDBsum e o programa PROCHECK [164]) e que manteve a abertura da suposta cavidade de ligação do substrato doador aberta. O sistema apo foi obtido a partir da remoção do peptídeo acceptor do arquivo PDB original da estrutura cristalográfica. Para a obtenção do complexo com os glicopeptídeos, a estrutura da glicana foi retirada do código PDB 2K33 [165]. O modelo do LLO foi obtido via o uso do método gromos de agrupamento (*clustering*) para que algumas conformações mais prevalentes e potencialmente relevantes fossem extraídas da DM, prontas para serem utilizadas em estudos de atracamento molecular (será detalhado mais adiante na seção de Métodos). O estado pós-catalítico foi obtido após a simulação do estado pré-catalítico, a partir da análise de componentes principais (ACP) que nos permitiu identificar regiões de menor energia do estado pré-catalítico, com um mínimo global e com mínimos locais.

A AglB possui estruturas completas obtidas pelo método de cristalografia por raios-X depositadas no banco de dados do PDB sob código 3WAK [66] e, mais recentemente (2017), 5GYM [78]. No início desse trabalho, a estrutura de código 5GYM na presença do peptídeo acceptor ainda não estava disponível, dessa forma não sendo considerada na elaboração desse capítulo. Na AglB, a EL5 está disponível na estrutura do 3WAK, porém não apresenta o íon sulfato que seria o equivalente ao fosfato do LLO, sendo a 3WAK uma estrutura apo. Os mutantes foram obtidos pela deleção da região P1 (resíduos 777 a 868) e da região IS (resíduos 638 a 776) do arquivo PDB da estrutura original.

A construção dos modelos de PglB quiméricos foram realizadas em duas abordagens. A primeira delas envolveu o uso do software MODELLER9v17 [166], utilizando como entrada um alinhamento de três sequências de aminoácidos utilizando o software Clustal Omega [167]: a sequência completa da PglB-quimérica que se quer modelar, a sequência da CIPglB e a sequência apenas da porção P1 da AfAglB-L. A rotina empregada na produção dos modelos foi a *automodel*. Vinte modelos foram construídos e avaliados estereoquimicamente pelo software PROCHECK no servidor PDBsum [164].

O modelo com a maior quantidade de resíduos em regiões permitidas foi selecionado.

4.2 Atracamento molecular

Duas abordagens distintas foram empregadas nos trabalhos dessa tese. No Capítulo I, buscando posicionar as glicanas corretamente na cavidade da PglB, elas foram atracadas rigidamente com auxílio do software AutoDock 4.2 [168]. O algoritmo genético Lamarckiano (LGA) foi empregado para explorar a cavidade de ligação. Em cada rodada, o número máximo de avaliações de energia empregado foi 250.000.000 e um número máximo de 27.000 operações de LGA foram gerados em populações de 10 indivíduos. No total, 100 rodadas foram realizadas, com valores de 0,80, 0,02 e 1 para cruzamento, mutação e elitismo. A estrutura de menor energia da glicana foi selecionada e então conectada à Asn aceptora ao adicionar os parâmetros relativos à ligação, ângulos e diedros na topologia do complexo e a subsequente realização de uma minimização de energia com o método de descida mais íngreme (*Steepest Descent*).

No Capítulo III, para a obtenção de um modelo do complexo pré-catalítico da OST, foi utilizado o servidor PatchDock [169], pois esse permite o uso de restrições espaciais em seus cálculos, como distâncias entre dois átomos, por exemplo. Conseqüentemente, baseado no conhecimento já descrito na literatura [158], as seguintes restrições foram aplicadas: i) distância menor que 4 Å entre o carbono 1 (C1) da bacilosamina e o átomo de nitrogênio (ND2) da amida na cadeia lateral do resíduo de Asn; ii) distância menor que 6 Å entre o primeiro fósforo do LLO (P1, mais próximo da cauda hidrofóbica) e os nitrogênios da cadeia lateral do resíduo Arg375 (NH1 e NH2); iii) distância menor que 6 Å entre o segundo fósforo do LLO (P2, mais próximo do primeiro monossacarídeo) e o metal divalente do sítio ativo da PglB. A estrutura proteica utilizada no atracamento foi a da PglB, sob código PDB 3RCE, na ausência da porção N-terminal da EL5. Ainda, a porção lipídica foi reduzida a apenas 3 unidades isoprenóides, possibilitando a redução do impedimento estérico no sulco hidrofóbico em um primeiro momento. Com esses parâmetros de entrada foi possível obter um modelo da OST bacteriana com todos os substratos. Para obter a porção N-terminal faltante da EL5 nesse modelo, foram recuperados os estados conformacionais identificados nas simulações vistas no artigo do Capítulo I (Figura 4, painel inferior). Quanto ao restante da cauda hidrofóbica do LLO, realizamos um alinhamento da sua estrutura original com a estrutura reduzida utilizada para o atracamento. Por fim, após a realização de minimizações de energia, os contatos impeditivos foram removidos e o sistema permaneceu estável para iniciarmos os estudos do complexo.

4.3 Inserção em membrana

As OSTs são proteínas transmembranas, por conseguinte, para melhor reproduzir as condições encontradas nos organismos, o emprego de membranas biológicas no desenvolvimento dos sistemas estudados faz-se necessário. A PglB foi inserida em uma membrana de 1-palmitoil-2-oleil-sn-glicero-3-fosfoetanolamina (POPE) com 512 fosfolipídios. A metodologia utilizada para a inserção da enzima da maneira mais apropriada foi a InflateGRO [170]. Ela consiste em sobrepor a estrutura da enzima e da membrana manualmente na posição onde se deseja inserir a estrutura e, subsequentemente, expandir as dimensões da caixa de simulação, ao passo que os lipídios são translacionados lateralmente na mesma proporção, multiplicando essas dimensões por um fator de escalonamento comum. Posteriormente, os fosfolipídios são lentamente redirecionados para o centro da caixa, concomitantemente à redução das dimensões da caixa de simulação, enquanto uma série de minimizações de energia é realizada em cada iteração, até que a área por lipídio atinja valores próximos ao inicial (que devem estar de acordo com dados experimentais). 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. Como forma de guiar a inserção, foi respeitada uma distância estimada de 8 Å na posição do íon catalítico Mg^{2+} em relação à membrana.

A AglB também foi inserida em uma membrana de POPE com 512 fosfolipídios. A metodologia utilizada para a inserção da proteína da maneira mais apropriada foram as ferramentas InflateGRO 2 e LAMBADA [171], uma combinação que representa uma evolução do método InflateGRO. O algoritmo LAMBADA calcula um escore do padrão de hidrofobicidade da proteína para otimizar o alinhamento com a membrana automaticamente. O InflateGRO2 representa um melhoramento do seu antecessor, realizando o processo de forma automática, reduzindo a quantidade de expansão, compressão e minimização da membrana, mantendo o empacotamento lipídico e reduzindo a exposição dos mesmos ao vácuo. Assim, a membrana biológica pré-equilibrada é mantida o mais próximo possível do seu estágio inicial. Por fim, os sistemas finais foram simulados com membranas possuindo 510 moléculas de POPE.

4.4 Dinâmica molecular

Os pacotes de simulações GROMACS versão 4.5.1 [172] e GROMACS 5.1.1 foram utilizados para a realização das DMs. Para a descrição dos parâmetros físico-químicos

das moléculas estudadas, os campos de força GROMOS 53A6 (Capítulo I) [173] e GROMOS 54a7 (Capítulos II e III) foram empregados, em conjunto com os parâmetros do GROMOS 53A6GLYC para a parte sacarídica (Capítulos I e III) [174, 175]. Os modelos e os parâmetros da membrana POPE foram retirados do repositório LipidBook [176], onde os dados de simulações de membranas variadas podem ser armazenados. Nesse caso específico, foi utilizado o trabalho de Piggot e colaboradores [177], empregando como base alguns parâmetros já descritos no GROMOS53a6. O modelo de água utilizado na solvatação de caixas triclinicas foi o SPC [178], na presença de condições periódicas de contorno. As minimizações de energia foram conduzidas no algoritmo Steepest Descent. O tratamento eletrostático escolhido foi o método PME [179]. Os comprimentos das ligações covalentes foram fixados pelo método de Lincs [180], permitindo a aplicação de tempos de integração de 2 fs. Nos Capítulos I e II, o esquema de *cutoff* era o *Group*, onde foram empregados valores de 1,2 nm para as interações coulômbicas e 1,4 nm para as interações de van der Waals. No Capítulo III, com o uso do GROMACS 5.1.1., um novo esquema de *cutoff* foi empregado por ser mais eficiente, onde os valores para as duas interações são necessariamente iguais (1.2 nm). Para evitar artefatos, foram utilizadas, além do tratamento por PME, correções de dispersão para os termos de van der Waals.

As simulações por DM, em geral, são divididas em duas fases: a fase de equilíbrio e a fase de produção. Na equilibração, as biomoléculas de estudo têm imposta sobre si uma restrição de posição para que seja possível uma acomodação gradual da orientação dos outros componentes do sistema, como as moléculas de solvente e os contraíons. Nos trabalhos dessa tese envolvendo as OSTs, duas simulações foram realizadas nessa primeira fase, empregando uma constante de força 1000 kJ/mol como penalidade energética para impedir o movimento. A primeira simulação é realizada com *ensemble* canônico (volume e temperatura constantes - NVT) visando a um ajuste adequado da temperatura do sistema. A segunda simulação é realizada com *ensemble* isobárico-isotérmico (NPT) de forma que a pressão e a densidade corretas da caixa sejam atingidas. Os tempos de simulação dessa fase variaram conforme o capítulo:

- Capítulo I: 0.1 ns (NVT) e 0.9 ns (NPT);
- Capítulos II e III: 1 ns (NVT) e 20 ns (NPT).

Após esse etapa, onde ocorre a equilibração dos fosfolípidios ao redor da proteína transmembrana, bem como do solvente, é iniciada a fase de produção (também NPT), quando ocorrerá a coleta dos resultados. Nessa fase, o sistemas completos (OSTs, LLO, membrana, solvente e íons) têm seu movimento irrestrito, sendo removida a restrição

de posição anteriormente imposta, o que permite a avaliação da dinâmica do sistema. Nesse período, novamente, os tempos de simulação empregados variaram conforme o Capítulo:

- Capítulo I: 200 ns, com o sistema PglB-Peptídeo em triplicatas;
- Capítulo II: 300 ns, todos sistemas em triplicatas;
- Capítulo III: 1 μ s para todos sistemas.

Na fase NVT, o acoplamento de temperatura selecionado foi o termostato V-rescale [181]. Na fase de NPT, foi empregado o termostato Nosé-Hoover [182, 183], pela sua característica de permitir maiores oscilações no acoplamento, o que auxilia na melhor descrição do movimento natural de sistemas envolvendo membranas. O acoplamento de pressão empregado foi o barostato Parrinello-Rahman [184, 185], pois reproduz a termodinâmica do sistema de forma apropriada, em combinação com o Nosé-Hoover, produzindo um *ensemble* NPT verdadeiro. A temperatura simulada para os sistemas bacterianos (PglB e LLO) foi 310 K, dentro da temperatura ótima de crescimento de *C. lari* [186], enquanto os sistemas de arqueas e as PglBs quiméricas foram simulados em 356 K, a temperatura ótima de *A. fulgidus* [187].

Nas simulações do fragmento de geraniol, novamente o pacote GROMACS 4.5.1 e o campo de força GROMOS 54a7 foram empregados. As simulações de equilíbrio foram feitas com algoritmo de acoplamento de Berendsen [188] por relaxar o sistema de forma eficiente. A simulação de produção da fase líquida foi calculada por 100 ns após a equilíbrio, com tempo de integração de 2 fs e o integrador leapfrog [189]. Para tratamento eletrostático, o método PME [179] foi utilizado, com uma *switching distance* de 1,1 nm. O algoritmo de acoplamento Nosé-Hoover foi utilizado como termostato, já que produz flutuações adequadas, importante aspecto para o cálculo de propriedades físico-químicas flutuantes [182, 183], com constante de acoplamento de 1 ps. Nas simulações do tipo NPT, o algoritmo de acoplamento Parrinello-Rahman foi utilizado como barostato, aplicando uma constante de 5 ps. Na simulação do fragmento em fase gasosa, uma única molécula foi simulada no vácuo durante 100 ns. O algoritmo LINCS foi utilizado, bem como o integrador SD, já que esse supera erros de graus de liberdade rotacionais e translacionais no vácuo [190].

4.5 Metadinâmica

Os cálculos de metadinâmica [152], uma forma de dinâmica molecular de amostragem ampliada, foram executados para os dissacarídeos isoladamente e para a unidade Asn-Bac, empregando os ângulos diedrais das ligações glicosídicas (ϕ e ψ) como variáveis coletivas (*collective variable* - CV). Essas CVs são submetidas a um viés (*bias factor*) que estimula essas características a popularem diferentes estados. Cada simulação foi realizada por 10 ns, empregando uma Gaussiana com altura de 1,2 e um σ de 0,35 para cada ligação glicosídica ϕ e ψ . O pacote GROMACS 4.5.1 em conjunto com o pacote PLUMED 1.2.2 [191] foi empregado para realizar as simulações. As superfícies de energia livre para os pares de ângulos foram obtidas aplicando o comando `sum_hills`, presente no pacote PLUMED.

4.6 Cálculos de mecânica quântica para a coordenação da PglB

Os resíduos catalíticos (Asp56, Arg147, Asp154, Asp156 e Glu319), juntamente com a Asn do peptídeo acceptor, foram submetidos a cálculos utilizando a Teoria de Funcional da Densidade (*Density Functional Theory* - DFT) [192] na base B3LYP/6-31G*. A geometria cristalográfica sem águas e três geometrias representativas das simulações por DM serviram de ponto de partida para os cálculos. Foram incluídas as cadeias laterais dos resíduos, enquanto o esqueleto amino acídico foi simplificado para um grupo metila no carbono β . Os cálculos *ab initio* foram realizados com o software GAMESS [193]. As geometrias tiveram sua energia minimizada com cálculos na base 3-21. Essas conformações foram ao cálculo DFT, permitindo a análise por matrizes de Hessian para caracterizar as conformações obtidas como mínimos da superfície de energia potencial.

4.7 Parametrização do LLO

Em colaboração com o doutorando Pablo Ricardo Arantes do Grupo de Bioinformática Estrutural realizamos a modelagem de um LLO de *C. lari* (Figura 13). Esse modelo foi obtido empregando parâmetros pré-existentes no campo de força GROMOS53a6 para a cauda hidrofóbica e a cabeça polar do lipídeo. Para a porção glicosídica, utilizamos o conjunto de parâmetros do GROMOS53a6GLYC [174, 175], compatível com o GROMOS54a7 [194]. No entanto, um parâmetro diedral presente nas unidades isoprenóides (destacado na Figura 13) da cauda hidrofóbica não existia no GROMOS54a7. Portanto, foi necessário a geração de um novo potencial baseado em fragmentos de

geraniol por cálculos quânticos. Para tal fim, o programa MOLLEN [195] foi utilizado na modelagem do fragmento, enquanto o programa GAMESS [193] foi empregado no cálculo do perfil torsional quântico desse diedro. Essa torção foi orientado em 0° e rotacionada em passos de 30° até completar uma rotação completa de 360° , tendo sua energia total calculada com o método HF e a base 6-31G*. Esse método de rotação foi utilizado de forma análoga para a construção do perfil torsional na MM, com a orientação dos ângulos diedrais sendo mantidas sob restrição de posição de 1000 kJ/mol. Os perfis advindos dos dois métodos foram submetidos ao servidor RotProf [196] para comparação e obtenção de novos parâmetros da MM que possuem padrão similar ao visto na QM.

Buscando a validação desse parâmetro diedral obtido, foram necessárias simulações em fase líquida e gasosa do fragmento de geraniol para o cálculo da entalpia de vaporização (ΔH_{vap}), conforme executado previamente [190]. Para construir a caixa de simulação do geraniol, uma única caixa cúbica de 2 nm de aresta foi criada contendo uma única molécula. Essa caixa foi replicada e empilhada de modo a gerar 125 novas caixas, cada uma contendo uma molécula de geraniol, formando uma caixa maior. Essa, por sua vez, foi submetida a uma simulação por DM em alta pressão (100 bar) para induzir as moléculas a atingirem a fase líquida, sendo, subsequentemente, realizada uma segunda simulação, agora em pressão 1 bar para equilíbrio do sistema. Após essa primeira fase, o sistema foi escalonado novamente (2x2x2), atingindo a quantidade de 1000 moléculas e simulada por mais 100 ns para coleta dos dados. Na fase gasosa, a simulação foi realizada com uma única molécula no vácuo por 100 ns. Esses dados foram utilizados para o cálculo de dois parâmetros comparativos: densidade (ρ) e entalpia de vaporização (ΔH_{vap}). O cálculo de ρ em uma simulação a pressão constante é regido pela massa do sistema M dividida pelo seu volume total V ($\rho = M/V$). A ΔH_{vap} pode ser obtida a partir da equação: $\Delta H_{\text{vap}} = (E_{\text{pot(g)}} + k_B \cdot T) - E_{\text{pot(l)}}$, sendo $E_{\text{pot(g)}}$ = a energia potencial da fase gasosa, $E_{\text{pot(l)}}$ = a energia potencial da fase líquida, k_B = constante de Boltzmann e T = temperatura (298 K).

4.8 Reconstrução filogenética

Para realizar a inferência das relações evolutivas das OSTs sequências de aminoácidos presentes no banco de dados do National Center for Biotechnology Information [197] foram buscadas. Uma abordagem mista foi empregada para desempenhar tal tarefa: buscas no servidor Basic Local Alignment Search Tool (BLAST) [198], utilizando a sequência de *C. lari* como entrada; busca manual por palavras-chaves "PglB", "AglB", "Stt3" e "oligosaccharyltransferase", sempre checando a presença da assinatura conservada

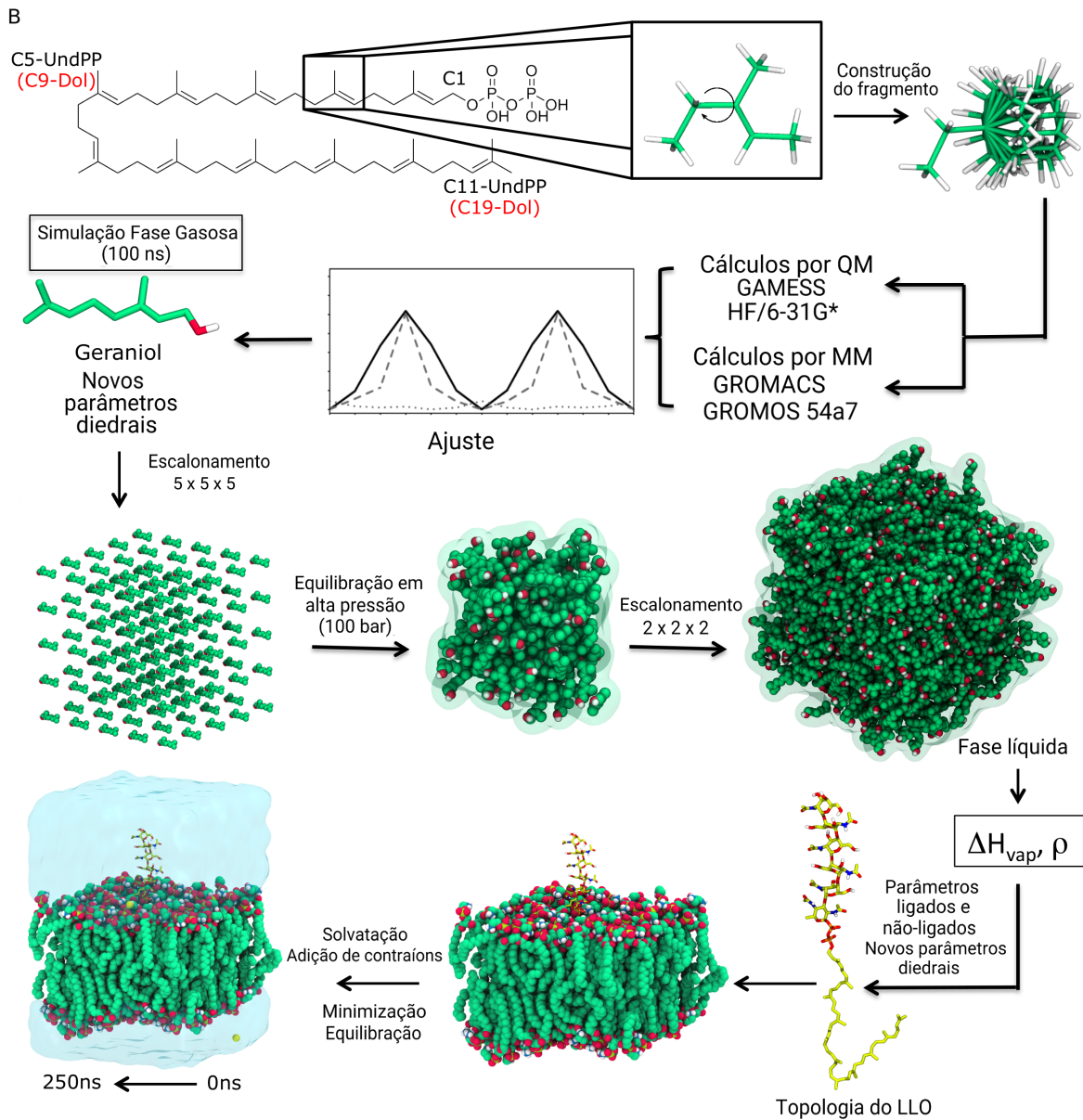


Figura 13 – Fluxograma representativo dos passos realizados para a modelagem do LLO de *C. lari*.

WWDXGX. Após essa varredura, um total de 210 sequências foram obtidas para os três grandes domínios. O alinhamento das sequências foi realizado com o algoritmo ClustalOmega [167], que permitiu o uso de um perfil de HMM (*Hidden Markov-Models*) extraído do banco de dados Pfam [199] relativo à OSTs.

O melhor modelo de substituição foi definido como o de Whelan and Goldman (WAG) [200] utilizando o ProtTest [201], incorporando a frequência média de aminoácidos dos dados utilizados (+F). Ainda, utilizou-se a distribuição Gamma para lidar com as diferentes taxas de evolução de cada sítio (cinco categorias, +G), bem como a prevalência de sítios invariantes evolutivamente (+I). Com o software MEGACC [202], uma árvore filogenética foi calculada pelo método de Máxima Verossimilhança (MV), aplicando um valor de bootstrap (replicatas) de 1.000. Para a manipulação e edição da árvore, o software FigTree 1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) foi utilizado.

4.9 Análises

Diversas propriedades podem ser avaliadas quando uma simulação por DM é gerada. Alguns exemplos são propriedades como volume, densidade, energia potencial, distâncias entre átomos, contatos nativos, entre outros. De maneira geral, a melhor forma de escolher as análises que serão empregadas está relacionada aos objetos de estudo, caso a caso. Entretanto, algumas análises são mais frequentemente encontradas, por isso serão explicadas com mais detalhes nessa tese.

4.9.1 RMSD

A análise do desvio quadrático médio (*root mean square deviation* - RMSD) é amplamente utilizada na DM para comparar a variação estrutural entre as estruturas da DM e uma estrutura de referência (em geral, a inicial). Essa é uma propriedade que pode ser observada ao longo do tempo, indicando alterações na estrutura estudada conforme a DM avança. A comparação é feita entre pares de estruturas, aplicando a seguinte equação:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (4.1)$$

Onde N é o número de átomos e δ corresponde a distância entre o átomo i e uma estrutura de referência ou a posição média de N átomos equivalentes.

4.9.2 RMSF

A análise da flutuação quadrática média funciona como medida da flexibilidade do sistema, podendo ser demonstrada em valores por átomos ou por resíduos. Diferentemente do RMSD, o dado gerado é um valor médio para toda a simulação e demonstra a flutuação da posição da estrutura investigada. A seguinte equação rege essa análise:

$$RMSF_i = \sqrt{\frac{1}{M} \sum_{k=1}^M (r_i(t_k) - \tilde{r}_i)^2}, \quad (4.2)$$

Onde M é o número de frames da simulação, $r_i(t_k)$ corresponde a partícula i do complexo r em um tempo k, e \tilde{r} corresponde a referência.

4.9.3 PCA

A análise de componentes principais (*principal component analysis*, PCA) consiste em uma forma de reduzir a dimensionalidade dos movimentos moleculares obtidos na DM, transformando-os em alguns poucos modos ou componentes importantes para a descrição do sistema [145]. O PCA permite a extração de componentes contendo a maior variação do sistema, ao passo que expressa essa informação como novas variáveis ortogonais chamadas de componentes principais [203]. No caso da DM, a informação é a variação de posição entre os átomos, geralmente as coordenadas cartesianas dos carbonos-. Utilizando combinações lineares desses dados, os componentes principais são obtidos, onde o primeiro componente principal (PC1) é aquele que contém a maior variação (e, possivelmente, a mais relevante) dos átomos do sistema [203]. Por conseguinte, via de regra, os dois primeiros principais componentes são os que descrevem a maior parte das conformações encontradas no conjunto de dados estudado. A partir deles é possível gerar gráficos bidimensionais em que as projeções dos PCs demonstram um comportamento conformacional da molécula de interesse, como um caminho percorrido durante a DM.

5 Resultados

*“Quando nós bebemos café,
as ideias marcham como um exército”*

Honoré Balzac

Os resultados da presente tese foram organizados nos seguintes capítulos, conforme indicado nos objetivos:

I. Descrição da dinâmica do sítio catalítico da PglB de *C. lari*;

Conrado Pedebos, Pablo R. Arantes, Guilherme M. Giesel e Hugo Verli: *In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology*, 2015, 25 (11), 1183-1195.

Esse trabalho demonstra as primeiras simulações por DM de uma OST, detalhando o comportamento do sítio ativo e a coordenação do metal da CIPglB com técnicas de MM e QM.

II. Avaliação do papel funcional das unidades estruturais da AglB de *A. fulgidus*;

Conrado Pedebos & Hugo Verli: Assessment of the Deletion of Structural Units in the Oligosaccharyltransferase AglB.

Nesse Capítulo, redigido em formato de um manuscrito pronto para submissão, é discutido como as unidades estruturais que compõem o domínio C-terminal da AfAglB-L podem influenciar na manutenção de sua estabilidade em altas temperaturas.

III. Caracterização dos estados pré- e pós-catalíticos da PglB de *C. lari*;

Essa parte da tese traz dados relativos à parametrização do LLO e à realização de simulações por DM envolvendo a CIPglB contendo tanto o peptídeo acceptor como o LLO, formando assim o complexo completo da OST bacteriana.

5.1 Capítulo I

Nesse Capítulo são descritas as primeiras simulações por DM envolvendo a CIPglB inserida em uma membrana biológica de POPE na presença de solvente explícito e os substratos proteico e glicoproteico. Essa linha de pesquisa foi iniciada durante meu Mestrado [159], entretanto, percebemos que havia necessidade de ampliar as investigações devido à complexidade dos sistemas. Durante o Doutorado focamos principalmente na descrição da geometria de coordenação do sítio ativo, empregando métodos de QM para esse fim. Realizamos ainda mais duas replicatas para o sistema da PglB-Peptídeo, buscando aumentar a amostragem das interações entre os resíduos do sítio catalítico e o peptídeo acceptor, o que nos possibilitou verificar novos estados conformacionais para a Asn acceptora.

Esse trabalho [160] aborda, primeiramente, aspectos estruturais gerais, observando a influência dos substratos ou de sua ausência na estrutura global da PglB. Concomitantemente, foi estudado o comportamento da EL5 na presença desses substratos distintos, possibilitando a identificação de alguns movimentos, em especial no domínio PP, que parecem ser relevantes para o acoplamento do acceptor, bem como para a liberação do produto. Posteriormente, uma detalhada análise do sítio ativo e seus entornos foi realizada, caracterizando a geometria de coordenação do íon metálico, o comportamento da Asn do acceptor e descrevendo interações importantes entre os resíduos que compõem as cavidades onde ocorre a catálise.

Computational Biology

***In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination**

Conrado Pedebos, Pablo Ricardo Arantes, Guilherme Menegon Giesel, and Hugo Verli¹

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

¹To whom correspondence should be addressed: Tel: +55-51-3308-7770; Fax: +55-51-3308-7309; e-mail: hverli@cbiot.ufrgs.br

Received 21 January 2015; Revised 13 July 2015; Accepted 15 July 2015

Abstract

The last step of the bacterial N-glycosylation pathway involves PglB, an oligosaccharyltransferase, which is responsible for the *en bloc* transfer of a fully assembled oligosaccharide chain to a protein possessing the extended motif D/E-X-N-X-S/T. Recently, this molecule had its full structure elucidated, enabling the description of its domains and the proposition of a catalytic mechanism. By employing molecular dynamics simulations, we were able to evaluate structural aspects of PglB, suggesting prevalent motions that may bring insights into the mechanism of the glycosylated peptide detachment. Additionally, we identified transient states at the catalytic site, in which the previously described carboxamide twisting mechanism was observed. Aided by quantum mechanics calculations for each different conformational states of the catalytic site, we determined the presence of an octahedral metal coordination, along with the presence of one water molecule at the catalytic site.

Key words: computational glycobiology, enzyme catalysis, molecular dynamics, oligosaccharyltransferase, PglB

Introduction

N-glycosylation is among the most prevalent protein modifications in nature (Varki 1993). This process is found widespread in all domains of life and is mainly associated in Eukaryotes (Varki 1993) to signaling, stability, trafficking, adhesion, intracellular location and immune response, in Archaea (Abu-Qarn et al. 2007; Eichler 2013) to protection against harsh environments, and flagella assembly, and in Bacteria (Szymanski et al. 2002; Nothaft and Szymanski 2013) to host adhesion, colonization and invasion. In the latter, these posttranslational changes were not reported until 1999 (Szymanski et al. 1999). Regarding the N-glycosylation pathway in Bacteria, after the full assembly of the oligosaccharide chain, the catalysis of the *en bloc* transfer is performed by oligosaccharyltransferase (OST) enzyme PglB (Wacker et al. 2002) to a protein that has the motif D/E-X-N-X-S/T (Kowarik, Numao, et al. 2006; Kowarik, Young, et al. 2006).

In 2011, this biomolecule had its full structure characterized (Lizak et al. 2011) by X-ray crystallography, allowing a more comprehensive look at the complex array of interactions behind the mechanism of action concerning the oligosaccharide transfer. PglB (Figure 1) consists of a transmembrane (TM) domain, from residue 1 to 432, and a periplasmic (PP) domain, from residues 433 to 712. The TM portion comprises 13 helices inserted in the membrane and two external loops (EL), named EL1 and EL5. The latter is believed to play an important role in catalysis, since Glu319 is part of the catalytic site and may pin the acceptor protein at the binding site. The other two residues that compose the catalytic site are Asp56, which resides in EL1, and Asp154, in a short PP loop. In addition, a conserved residue, Arg331 (Figure 1—bottom right panel), was suggested as providing interactions with the extended glycosylation sequon (Asp/Glu in position -2) (Lizak et al. 2011; Gerber et al. 2013; Ollis et al. 2014).

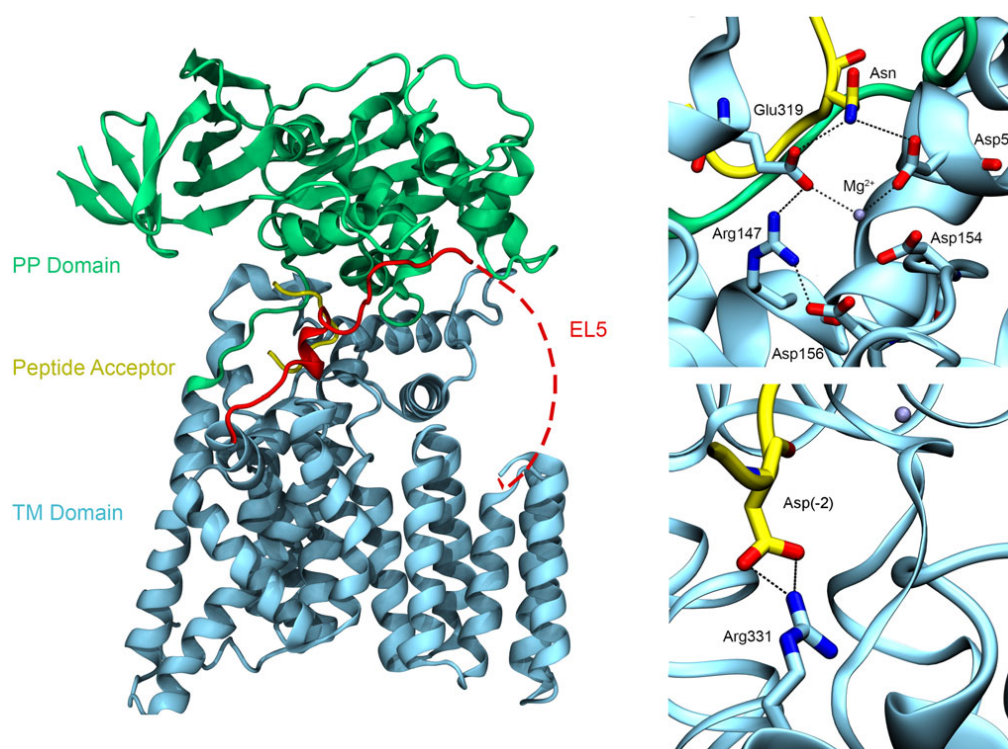


Fig. 1. PglB crystal structure and its interactions. (Top right panel) Catalytic site network of interactions, coordinating both the metal ion (Mg^{2+}) and the acceptor amide group from the Asn side chain. (Bottom right panel) Residue Arg331 interaction with Asp (-2) from the extended sequon. This figure is available in black and white in print and color at *Glycobiology* online.

As for the PP domain, PglB exhibits two different units, the central core (CC) and the insertion sequence (IS) unit (Igura et al. 2008; Maita et al. 2010). The CC unit presents the largely described motifs WWDYGY (residues 463–468) and MxxlxxxV (residues 569, 572 and 576) (Maita et al. 2010). 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 (Ser/Thr). Furthermore, Tyr468 from the first motif has been implied as responsible for interacting with the N-acetylation at the C2 position of bacillosamine (Lizak et al. 2011), after a tentative modeling of a substrate on its cavity. Additionally, two new motifs have been recently identified, the DxxK (residues 481–484), which is also implicated in the interactions with the glycan chain (Jaffee and Imperiali 2011) and the Tyr-plug (Tyr293 at the N-terminal half of EL5) (Lizak et al. 2014), which is essential for catalysis but not for protein binding. Finally, the IS structural unit is presented as a β -sandwich motif containing nine β -strands. Although its function is not yet determined, it is hypothesized that it is dispensable for catalysis (Matsumoto et al. 2012) and might be important for structural stability (Matsumoto, Shimada, Kohda, et al. 2013; Matsumoto, Shimada, Nyirenda, et al. 2013).

For the enzyme mechanism of action, two cavities are proposed: the protein acceptor binding site, which requires not only the consensus sequence but also a largely surface-exposed and flexible polipeptide region (Kowarik, Numao, et al. 2006; Kowarik, Young, et al. 2006), and a hydrophobic groove, where the hydrophobic portion of the lipid-linked oligosaccharide (LLO) would be positioned (Lizak et al. 2011), 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 (Figure 1—top right panel) along with the protein substrate and a divalent ion Mg^{2+} , which results in an optimal position and configuration of the nitrogen atom from Asn side chain to perform a nucleophilic attack on the C1 position from bacillosamine residue (Lizak et al. 2011). However, the achieved crystallographic resolution impaired an unequivocal determination of such coordination (Jaffee and Imperiali 2011). Still, recent biochemical evidences reinforced the hydrogen bond network (twisted carboxamide mechanism) by testing chemically synthesized acceptor peptide analogues (Lizak et al. 2013).

In our work, we employed a series of computational strategies intending to: (I) characterize the dynamics of PglB within the membrane environment; (II) include the substrate glycan chain searching for clues on substrate binding and unbinding; and (III) obtain atomic-level information on metal coordination and its role on catalysis, while in the presence of explicit water molecules.

Results

Systems setup

A total of four systems were studied: one simulation of PglB in its native uncomplexed state (PglB-Native), three replicate simulations of PglB bound to its peptide acceptor (PglB-Peptide I, II and III), one simulation for PglB bound to the glycopeptide found in *Campylobacter jejuni* (PglB-Glycopeptide *C. jejuni*), and one simulation for PglB bound to the glycopeptide found in *Campylobacter lari* (PglB-Glycopeptide *C. lari*). Each protein was inserted in a POPE membrane, employing the InflateGRO methodology (see the

Materials and methods section), followed by two equilibration steps (NVT and NPT). Finally, we performed six 200 ns production molecular dynamics (MD) simulations, which provided the collected data.

Global structural analysis of PglB

By analyzing the contribution of the enzyme domains for the total RMS (Figure 2), we observed that a lower TM domain value contributes to the lower total value of the PglB complexed with its substrates (all lines, except black), while the PP domain presented a more spread pattern. Concerning all peptide measures, the only differences were observed in the PglB-Glycopeptide *C. lari* oligosaccharide MD run (Figure 2, top right graph) and in one of the replicates of PglB-Peptide III, where the peptide rms undergoes an average shift of 0.2 nm.

Several regions demonstrated variations in their flexibility due to the presence of the substrates in the cavities (Figure 3). The PP domain portrays a high flexibility, while the TM helices present a more rigid pattern. The region that displayed the highest flexibility is the IS unit, which correlates to the distorted barrel shape. Three loops comprising residues 68–78 (a loop connecting helices EL1-h1 and EL1-h2), 283–306 (N-terminal half of EL5) and 516–526 demonstrated the highest structural fluctuations in all simulations. These structures are located circumventing the glycan binding site and move in different directions, depending on the analyzed simulation. In the PglB-Peptide simulations I and III, we observed diffuse movements of EL5, whereas not closing the cavity, while in simulation II, this structure is almost closing the glycan cavity (Figure 4), indicating different spatial orientations of the flexible N-terminal half of EL5 (residues 283–306). In the glycopeptide bounded states, these loops perform a movement that reduces the size of the glycan binding cavity,

while the same did not occur in the apo native state, in which the cavity remained open. Further analysis of the systems Free Energy Landscape as a function of the conformational ensemble produced by the simulations indicated that the native state presents a higher variability in its conformational profile, underlined by two different regions of minimal energy. The PglB-Peptide complex demonstrates two different patterns: one with a larger exploration of the conformational space and the other with less variation. In addition, both PglB-Glycopeptide systems suggest an increased rigidity, that is, a more restricted population of the conformational space when compared with the nonglycosylated systems.

The main motions related to the observed ensembles were depicted through examination of the conformations between the extremes projections from the first (the most representative) to the third principal components. In the PglB-Glycopeptide *C. jejuni* and *C. lari* simulations (Figure 5), the loops near the glycan cavity move toward the oligosaccharide chain, being in close contact with it. Concomitantly, the β -barrel-like IS unit appears to have an upward motion, which may expand the protein binding cavity for the glycopeptide departure (principal components—PC-1 and 2). Also, the IS unit perform a lateral movement (PC 3), increasing the area surrounding the protein site for the attachment of the acceptor substrate and release of the reaction's product.

Catalytic site behavior during molecular dynamics

Bacterial glycoproteins exhibit the extended motif D/E-X-N-X-S/T. We investigated two conserved residues, Arg331 and Arg147, for possible interactions with the Asp in position -2. Our data showed that Arg147 interacted for around 50% of total replicates time, while Arg331 interacted for around 8% of total replicates time, keeping

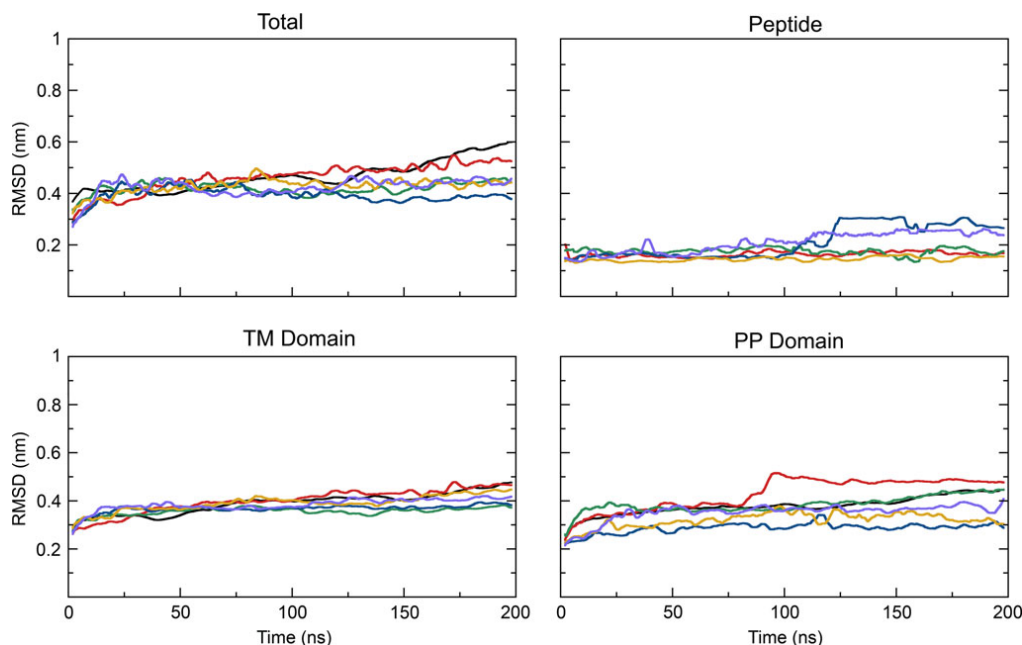


Fig. 2. RMSD analysis comprising the 200 ns MD simulations of the PglB. PglB-Native (black), PglB-Peptide replicates (orange—I, red—II and violet—III), PglB-Glycopeptide *C. jejuni* (green) and *C. lari* (blue). Results are shown for the total protein, for the bound peptide, for the TM domain and for the PP domain. Results are displayed after an averaging factor of 1000 points for visualization purposes. This figure is available in black and white in print and color at *Glycobiology* online.

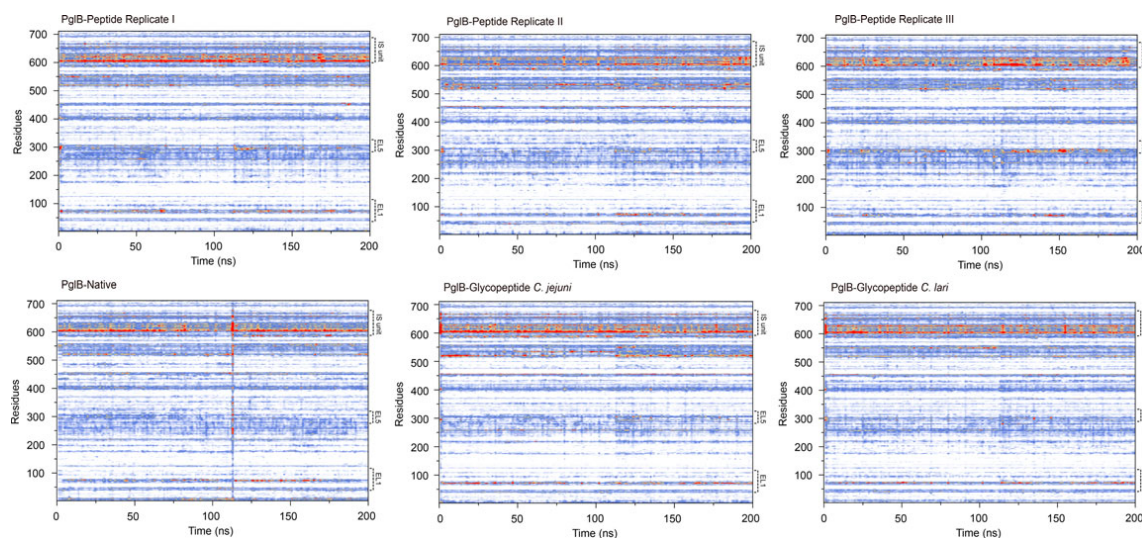


Fig. 3. RMSF heat map during the 200 ns simulations. The color range varies in a white–blue–orange–red heat map scheme (color version), changing from low structural fluctuation to high structural fluctuation. RMSF averages were collected at every 1 ns to construct the graphics. This figure is available in black and white in print and color at *Glycobiology* online.

the Asp (–2) residue in a tight binding with the acceptor site. As previously mentioned, due to the low crystal resolution, the hydrogen bond network at the catalytic site of PglB remains unresolved. The elucidation of this delicate arrangement is crucial for the comprehension of the enzyme’s catalysis mechanism. The three residues from PglB catalytic site are predicted to coordinate a divalent metal ion (Mg^{2+}). Consequently, we measured the distances between these atoms and the co-crystallized Mg^{2+} in the PglB-Peptide complex simulations I, II and III (pre-catalysis state), aiming to evaluate if they were compatible with metal coordination (Table I). Looking at the distances calculated in our models, the Mg^{2+} ion was firmly stable between residues in the performed simulations, as a result of interaction with six oxygen atoms presented, where each residue can contribute with either a monodentate and/or a bidentate coordination in an interchangeable mode.

The twisted carboxamide catalytic mechanism (Lizak et al. 2011) postulates that Asp56 and Glu319 residues would form hydrogen bonds with the terminal amide of the substrate Asn residue. This trait would permit an increase in the electronegativity of this group, modifying its geometry from the planar state to a 270° conformation, which causes the exposure of the lone pair of electrons. This would enable the amide group to perform a nucleophilic attack to the glycan anomeric carbon atom (Wiberg and Breneman 1992).

Observing the distribution of distances between Nδ2 of acceptor Asn and Asp56 Oδ1 and Oδ2 oxygen atoms, two distinct population peaks can be identified (Figures 6 and 7A): one corresponding to distance values compatible with hydrogen bonding, while the other does not show this feature. In PglB-Peptide I and III simulations, the first described peak was more populated, while PglB-Peptide II displays a higher population in the second one. This appears to be associated with a carboxamide position inversion that occurred in all simulations, only differing by the amount of time spent on this conformation. This transition happens by a reorientation of both the amide and the carbonyl dihedral (Figure 7B) angles ($C\beta-C\gamma-N\delta2-Mg^{2+}$ and $C\beta-C\gamma-O\delta1-Mg^{2+}$) from the Asn-acceptor side chain. This causes the amide to drift apart from the Mg^{2+} ion and the catalytic residues,

and to interact with the loop comprehending residues Asp481 to Gly487 (DxxK motif), specially with Gly482, Gly483 and His484 (Figure 7D—left panel). Such inversion seems to be supported by the crystal structure electron density, indicating an additional conformational state of the catalytic site (Figure 7D—right panel) that may be enzymatically inactive.

Another unresolved issue concerns the presence of water molecules participating in the coordination of the catalytic ion. Employing a radial distribution function (RDF) analysis (Figure 8A) to verify the presence of solvent atoms near the Mg^{2+} , we detected the presence of at least two hydration layers, in three patterns of coordination: the first one (Figure 8B) represents the amide nitrogen position of the crystal structure, while the second one (Figure 8C) is after the carboxamide inversion and occurs with the carbonyl coordinating the catalytic ion; the third one (Figure 8D) presents a water molecule occupying the position where the carboxamide previously was and, in addition, coordinates this group. Furthermore, in the second pattern, another water molecule is interacting with the carboxylate group of Glu319.

Catalytic site coordination evaluation by quantum mechanics

The three configurations observed for the Mg^{2+} ion, under the performed simulations, maintained an octahedral coordination, although molecular mechanics does not handle directly electronic properties. So, we employed a series of QM calculations aiming to characterize the atoms and interactions involved in Mg^{2+} binding (Tables II–IV). For such comparison, we selected the catalytic site crystallographic conformation, as well as the three MD extracted frames presented in Figure 8.

The octahedral coordination persisted on the three MD extracted frames during the QM calculations. The pattern of interactions with the six oxygen atoms was also maintained (Table II), demonstrating variations in the denticity of each carboxylate ligand. The crystallographic conformation displayed four atoms coordinating the ion,

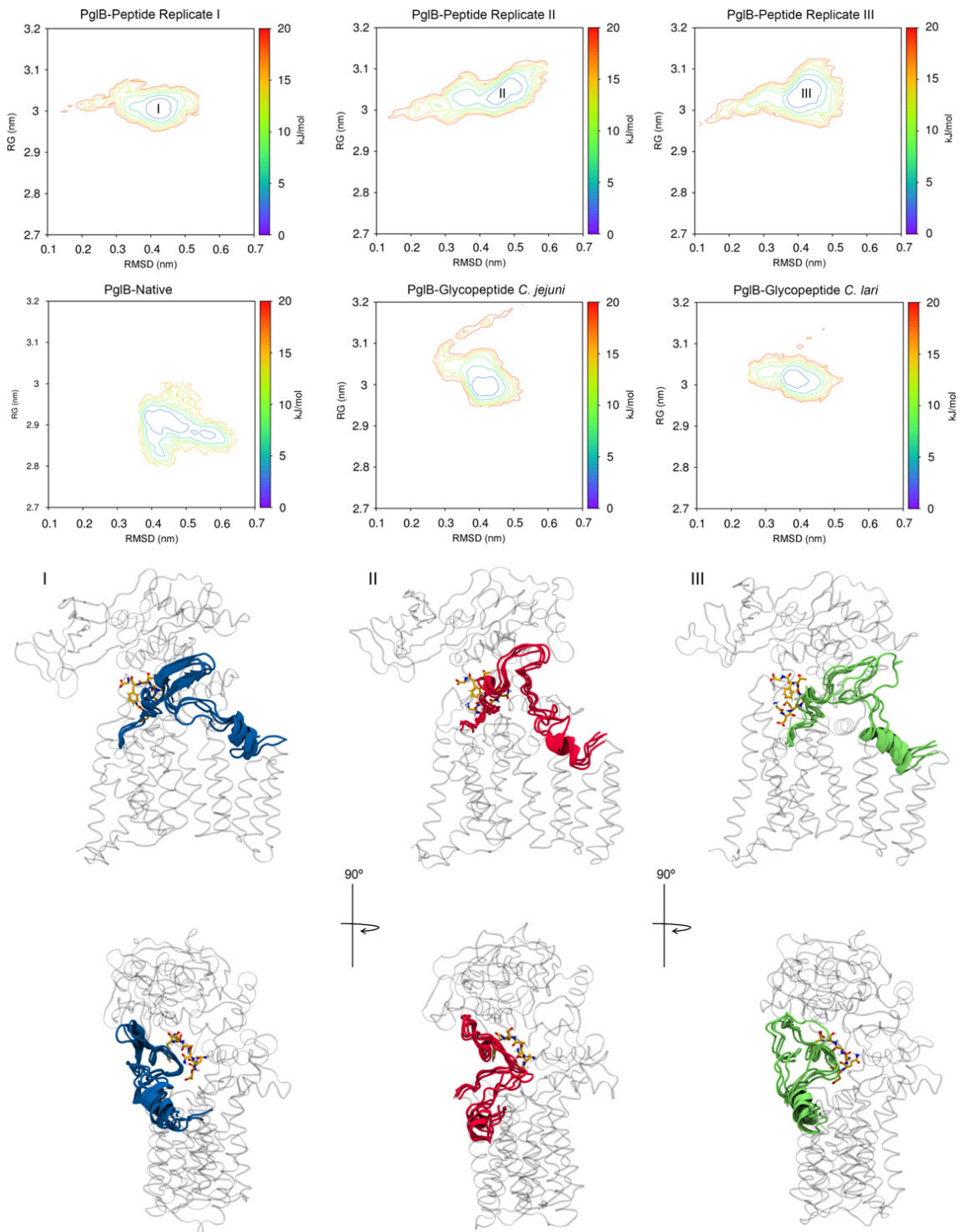


Fig. 4. Free energy landscapes and minimal energy structures extracted from PglB-Peptide simulations. (Top) Free energy landscapes represented as 2D contour plots for each MD run, based on the protein RMSD and radius of gyration (RG). (Bottom) The minimum energy representative structures from the PglB-Peptide simulations (I, II and III) are aligned, depicting the flexible behavior of EL5 (cartoon structures). Contour lines are shown at every 5 kJ/mol from 0 to 20 kJ/mol. This figure is available in black and white in print and color at *Glycobiology* online.

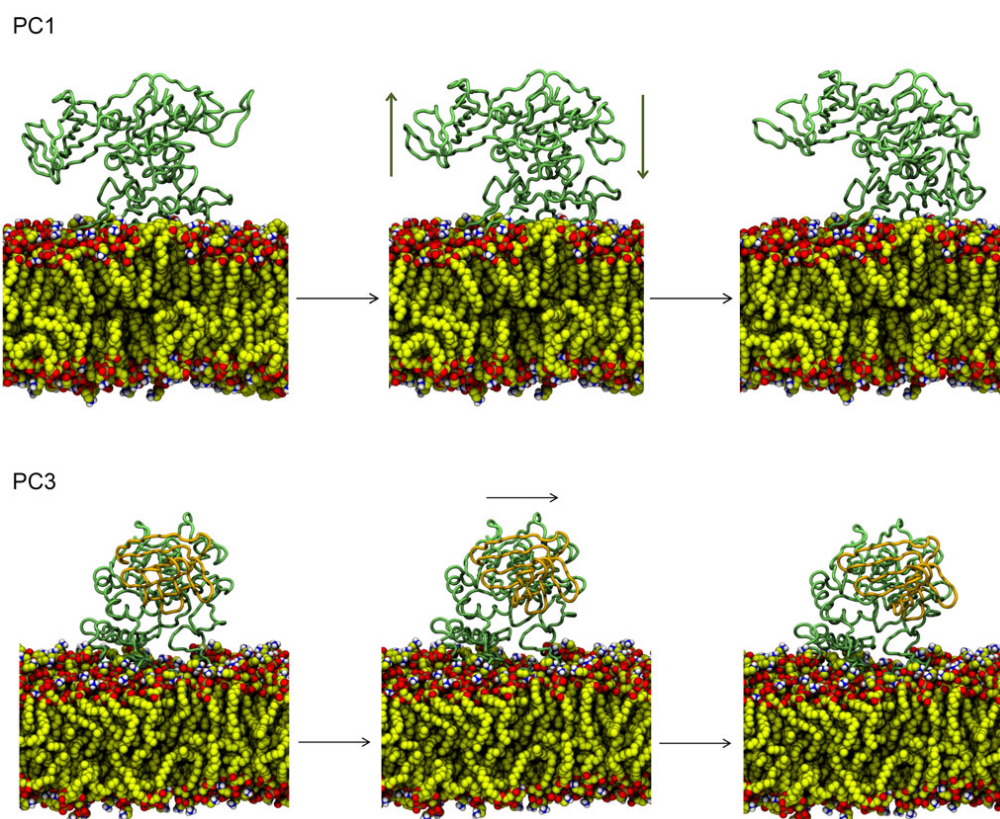


Fig. 5. PglB-Glycopeptide principal components analysis. Schematics of the transition extracted from the extreme points of the PC1 (front view) and the PC3 (side view, rotation of 90°) principal components. In PC3, the IS unit is highlighted. This figure is available in black and white in print and color at *Glycobiology* online.

Table I. Distances between the catalytic site ion Mg^{2+} and the side chains from the catalytic residues for each analyzed PglB-Peptide simulation (I, II and III)

Residue and atom	MD distance to Mg^{2+} (Å)		
	PglB-Peptide I	PglB-Peptide II	PglB-Peptide III
Asp56 Oδ1	2.14 ± 0.08	2.10 ± 0.10	2.09 ± 0.10
Asp56 Oδ2	2.07 ± 0.06	2.12 ± 0.08	2.13 ± 0.08
Asp154 Oδ1	2.08 ± 0.07	2.07 ± 0.06	2.09 ± 0.07
Asp154 Oδ2	2.08 ± 0.06	2.08 ± 0.06	2.08 ± 0.06
Glu319 Oε1	2.10 ± 0.07	2.12 ± 0.08	2.10 ± 0.06
Glu319 Oε2	2.08 ± 0.07	2.11 ± 0.07	2.12 ± 0.06

but there was no water molecule at the catalytic site, which lacked two additional ligands to achieve a coordination number of six (the most prevalent for Mg^{2+}) (Bock et al. 1999). Moreover, in the four frames, the Asn amide group interacted with one of the carboxylates (Asp56 and/or Glu319) through hydrogen bonds (Table III). To assess the twisting of the amide group, we calculated the dihedral angle Hδ1-Nδ2-Hδ2-Cγ, which showed values of 339.81° for the crystal, 346.26° for frame B', 306.36° for frame C' and 321.86° for frame D'. These results allowed us to identify a modification in the amide planar conformation (0 or 360°), with frame C exhibiting the highest deviation (−53.64 or 306.36).

Oligosaccharide conformational analysis

PglB has a promiscuous behavior, as the only prerequisite for the glycan substrate to be transferred is the N-acetylation at position C2 (Wacker et al. 2006). Consequently, we selected two oligosaccharide chains to simulate: the one from *C. jejuni*—composed by one Bacillosamine (Bac) residue, five N-acetylgalactosamine (GalNAc) residues and one glucose (Glc) residue—since it is the most studied system in Bacteria (Nothaft and Szymanski 2010) and previous NMR studies have characterized its structure (PDB ID: 2K33) (Slynko et al. 2009), as well as the one from *C. lari*, by removing the Glc residue ramification of the chain. Thus, by employing the oligosaccharide chain structure contained in the NMR structure from PDB ID 2K33, each glycosidic linkage from the PglB-Glycopeptides MD simulations was analyzed within protein cavities and compared with the glycopeptide in solution (Table IV). As demonstrated, the average conformational profile of the oligosaccharide chain attached to the peptide in the catalytic site remains mostly rigid during the simulations, with minor fluctuations to additional conformational states.

We expanded this comparison by performing metadynamics simulations of the minimal units (disaccharides and monosaccharide with Asn) from the glycan chain and analyzing the ϕ and ψ angles portraying the lowest energies (Figure 9). The obtained maps agreed with the geometries distribution from the MD simulations, except for the β -D-Bac-(1 → N)-Asn linkage, which in our MD simulations (PglB-Glycopeptide *C. jejuni*, PglB-Glycopeptide *C. lari*, and

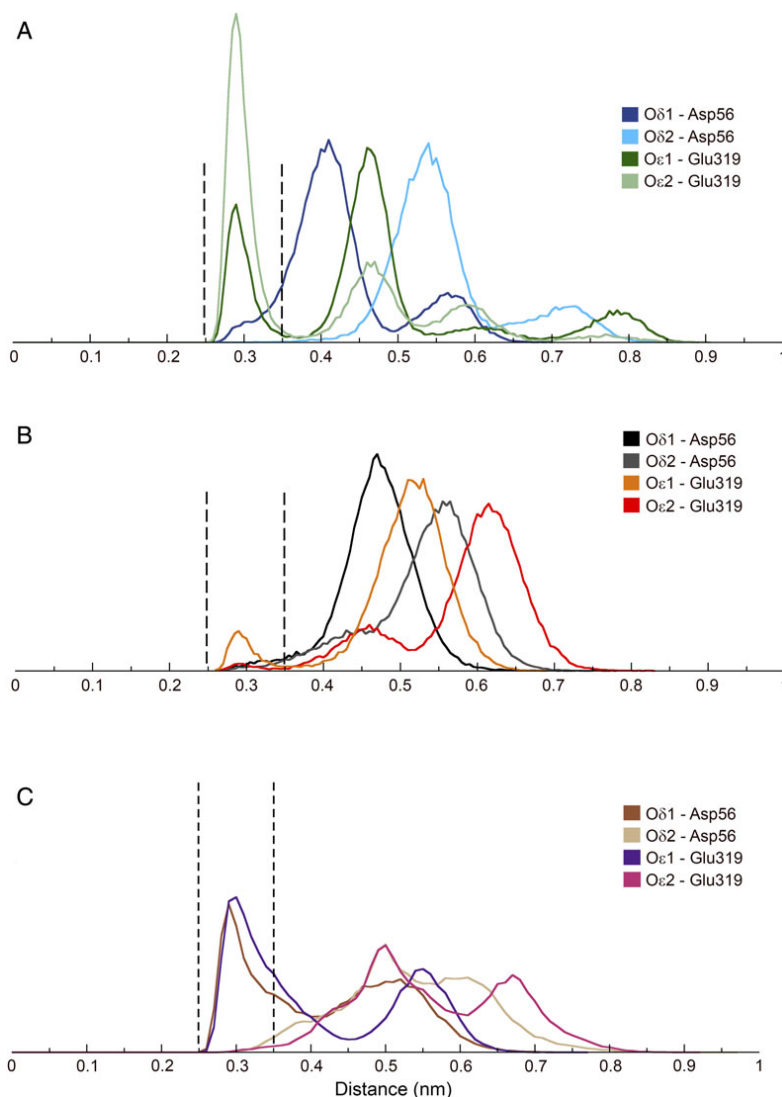


Fig. 6. Distance distribution (PglB-Peptide simulations I, II and III) between the catalytic residues side chain (Asp56 and Glu319) and the acceptor amide from the Asn side chain. (A) Asp56 and Glu319 from PglB-Peptide I. (B) Asp56 and Glu319 from PglB-Peptide II. (C) Asp56 and Glu319 from PglB-Peptide III. Highlighted with two dashed bars are the appropriate distances for hydrogen bonding (max. 0.35 nm). This figure is available in black and white in print and color at *Glycobiology* online.

uncomplexed Glycopeptide in solution) remained rigid and stable on a local minima, while a highly flexible behavior was seen in the energy contour plots. These data indicate that even outside the PglB binding cavity, the glycopeptide displays a similar distribution in solution for all the glycosidic linkages, that is, no major conformational induction appears to operate at the glycan chain.

Discussion

Although the crystal resolution of 3.4 Å, PglB structure maintained its stability throughout the simulations as indicated by its RMS, reinforcing the adequacy of the experimental information. The native uncomplexed state presented the highest RMS, which was expected, since the original crystal structure was obtained co-crystallized with its protein

substrate. In the complexed systems, the presence of these substrates appears to induce a more pronounced stabilizing effect on the protein structure, lowering the RMS total value. The fluctuations of RMS observed in the heat maps allowed the detection of high flexibility in the IS unit, and in the loops surrounding the glycan cavity site at all simulations. Concerning the former, the flexibility of that structure may facilitate the binding and unbinding of substrate and product, respectively, since a large and exposed flexible loop is necessary for the N-glycosylation (Kowarik, Numao, et al. 2006; Kowarik, Young, et al. 2006). As for the latter, the loop movements had different directions, as observed in Figure 4, promoting either exposure or closure of the glycan cavity site. Specifically, PglB-Peptide I, II and III simulations exhibited both behaviors, with the N-terminal half of EL5 adopting distinct conformations. In replicates I and III, the glycan

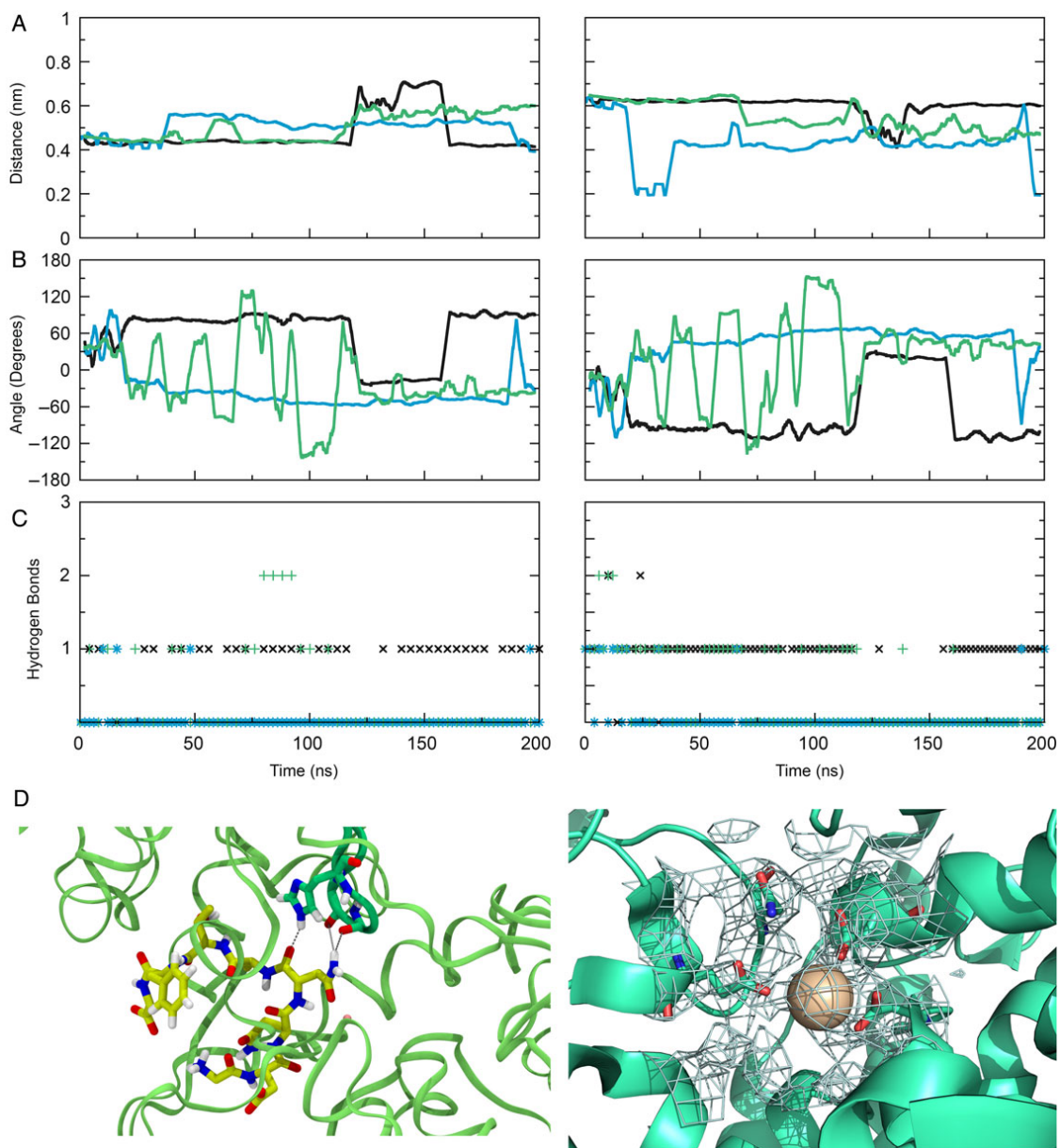


Fig. 7. Catalytic site interactions analysis. Black lines and X represent PglB-Peptide I; blue lines and * represent PglB-Peptide II; green lines and + represent PglB-Peptide III. (A) Distance of the Mg²⁺ ion regarding Asn carboxamide atoms Nδ2 (left graph) and Oδ1 (right graph). (B) Reorientation of the amide (left graph) and carbonyl (right graph) dihedrals during the 200 ns simulation. (C) Hydrogen bonds accounting for the side chains of Asp56 (left graph) and Glu319 (right graph), correlating with the atoms distances and dihedral transitions. (D) Interactions between the DxxK motif with the acceptor Asn side chain and backbone (left panel) and electronic density (right panel) extracted from the crystal structure. Results in (A), (B) and (C) are displayed after an averaging factor of 1000 points for visualization purposes. This figure is available in black and white in print and color at *Glycobiology* online.

cavity remained open, leaving enough space for oligosaccharide accommodation, while in replicate II that portion of the loop closed the cavity and appears to rearrange the presumed binding groove area for the lipid part of the LLO. Our findings reiterates that EL5 (and specially its N-terminal half) possess a remarkable plasticity, and that the presence of different ligands influences its dynamics by stabilizing different shapes of the LLO binding groove and cavities.

The large amount of interactions induced by the glycosylated peptide agrees with the steric tension hypothesis proposed by Lizak et al. (2011), although the ligand never exited the protein during our simulations. The identification of high flexibility in the loops near the glycan binding site, when the acceptor peptide was bound, demonstrated the influence of this ligand in the protein structure. As shown, the PP portion performs a twist and a lateral (IS unit) motion that may allow

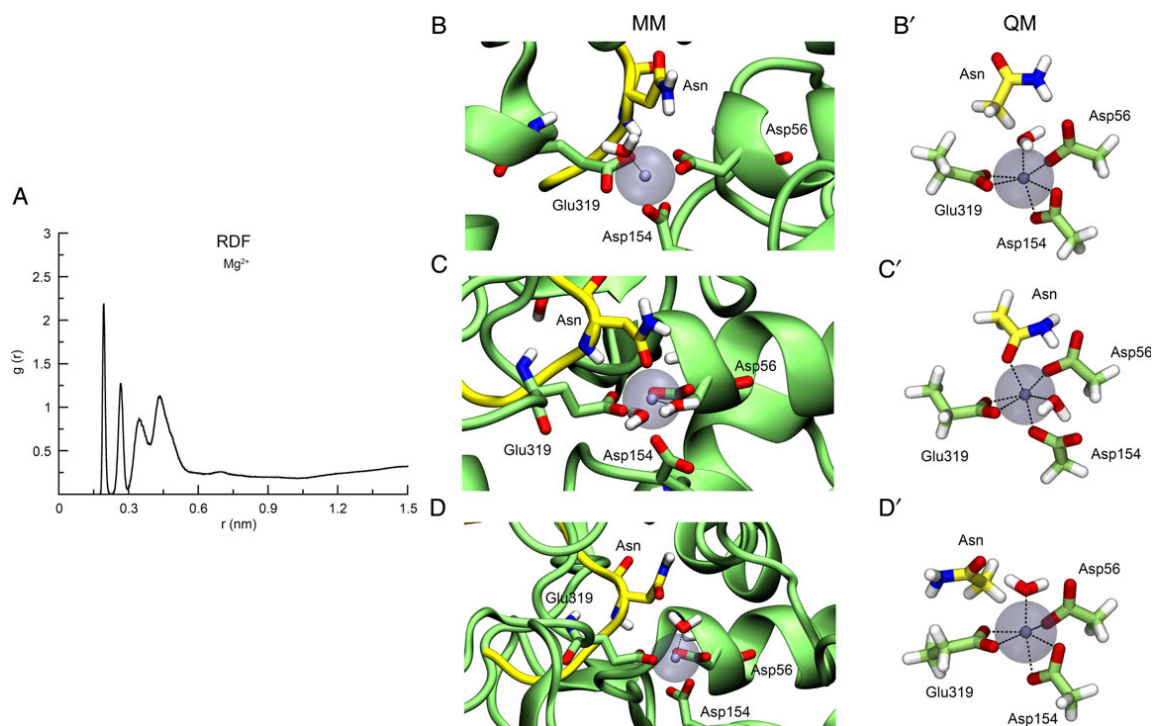


Fig. 8. Water molecule metal coordination. (A) RDF analysis demonstrating the presence of solvent molecules near the catalytic ion (<0.3 nm). (B–D) Snapshots extracted from the PglB–Peptide molecular mechanics simulations demonstrating the geometries of different interactions between the residues from the catalytic site with the divalent metal and the water molecules. (B'–D') Quantum mechanics evaluation final conformational states. All geometries of different interactions between the residues from the catalytic site with the divalent metal and the water molecule. This figure is available in black and white in print and color at *Glycobiology* online.

Table II. Distances and bond order between the catalytic site ion Mg^{2+} and the side chains from the catalytic residues for each analyzed frame on QM calculations

Residue and atom	Distance to Mg^{2+} (Å)		Bond order to Mg^{2+}	
	Crystal structure	Average of QM–MD frames	Crystal structure	Average of QM–MD frames
Asp56 Oδ1	4.14	3.59 ± 0.43	0	0
Asp56 Oδ2	1.94	2.02 ± 0.01	0.49	0.42 ± 0.01
Asp154 Oδ1	2.19	2.14 ± 0.13	0.42	0.33 ± 0.06
Asp154 Oδ2	2.12	2.53 ± 0.70	0.47	0.27 ± 0.23
Glu319 Oε1	3.95	2.20 ± 0.02	0	0.34 ± 0.03
Glu319 Oε2	2.11	2.13 ± 0.02	0.44	0.36 ± 0.03
Asn Oδ1	6.38	4.22 ± 2.08	0	0.10 ± 0.17
H ₂ O O	—	2.11 ± 0.02	—	0.38 ± 0.03

the expansion of the protein binding cavity, a movement that may be important not only to release the glycoprotein but also to better accommodate the protein binding at the acceptor cavity. In addition, the same motions may permit the compression of the oligosaccharide structure, causing steric tension at that region. Consequently, we suggest that an allosteric regulation may occur between sites, facilitating the entrance and the exit of the enzyme substrates.

Many PglB conserved residues had their roles previously predicted and/or described. Previous data (Gerber et al. 2013) employing mutagenesis examined the role of Arg331 in peptide binding and identified

Table III. Distance and bond order between Nδ2 and Oδ1 to C_γ atom in the carboxamide group of the asparagine residue, for each analyzed frame on QM calculations starting from different geometries (as shown in Figure 8)

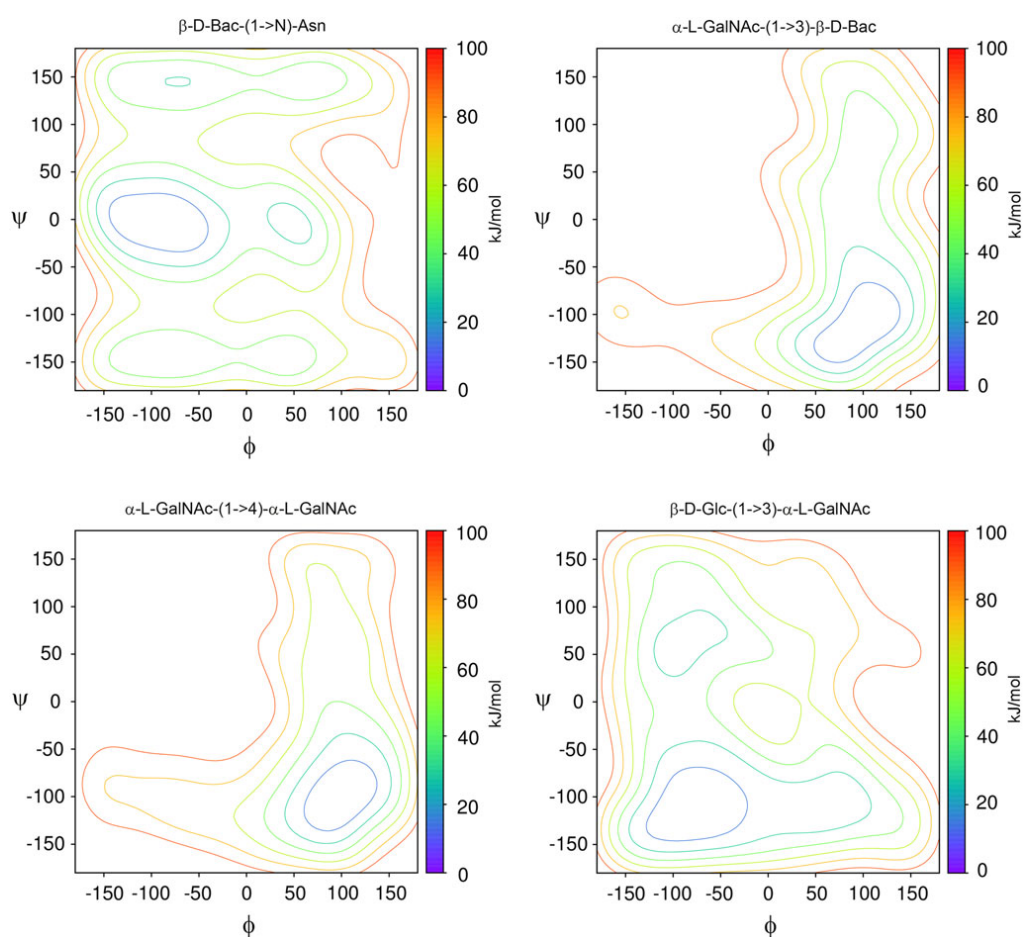
Residue and atom	Distance (Å)/Bond order to carbon of Asn			
	Crystal	Frame A'	Frame C'	Frame D'
Asn (Nδ2)	1.36/1.16	1.36/1.17	1.36/1.24	1.38/1.12
Asn (Oδ1)	1.23/1.83	1.23/1.82	1.24/1.57	1.23/1.74

that additional interactions were required for the binding of Asp (–2). The proposed candidate for this function was Arg147, previously predicted to participate in the catalysis network of interactions (Lizak et al. 2011). During the simulations, Arg147 interacted with Asp (–2) for half of the total simulation time, which was more than 6-fold the interaction time of Arg331. For this reason, we conclude that both Arginine residues contribute to peptide binding, a trait that explains the higher specificity of the extension of the motif. Besides, we quantitatively characterized the pronounced role of Arg147 in this process, with that being the first atomistic description of this behavior.

The observed distances for the carboxylate groups of the catalytic residues are compatible with metal coordination and are in closer agreement with experimental data (Katz et al. 1996) (2.073 Å) than the crystal structure. As previously discussed (Jaffee and Imperiali 2011), the low crystal resolution, together with the high pH of crystallization (9.4), may interfere with the proposed mechanism involving

Table IV. Dihedral angles of the glycosidic linkages composing the glycan obtained from the 200 ns MD of the glycopeptide in solution simulation

Glycosidic linkage	Uncomplexed glycopeptide (<i>C. jejuni</i>)		PglB-Glycopeptide (<i>C. jejuni</i>)		PglB-Glycopeptide (<i>C. lari</i>)	
	Dihedral angle (°)					
	ϕ	ψ	ϕ	ψ	ϕ	ψ
β -D-Bac-(1 \rightarrow N)-Asn	114 \pm 10	-141 \pm 11	111 \pm 9	-152 \pm 34	112 \pm 8	91 \pm 147
α -L-GalNAc1-(1 \rightarrow 3)- β -D-Bac	74 \pm 10	-98 \pm 14	74 \pm 7	-137 \pm 7	73 \pm 7	-135 \pm 7
α -L-GalNAc2-(1 \rightarrow 4)- α -L-GalNAc1	93 \pm 12	-100 \pm 15	74 \pm 9	-127 \pm 12	79 \pm 10	-119 \pm 14
α -L-GalNAc3-(1 \rightarrow 4)- α -L-GalNAc2	93 \pm 13	-93 \pm 8	97 \pm 8	-93 \pm 8	93 \pm 9	-96 \pm 9
α -L-GalNAc4-(1 \rightarrow 4)- α -L-GalNAc3	90 \pm 15	-109 \pm 19	68 \pm 10	-148 \pm 17	103 \pm 12	-86 \pm 11
α -L-GalNAc5-(1 \rightarrow 4)- α -L-GalNAc4	93 \pm 11	-98 \pm 13	92 \pm 9	-95 \pm 9	95 \pm 10	-94 \pm 11
β -D-Glc-(1 \rightarrow 3)- α -L-GalNAc3	-61 \pm 13	-117 \pm 20	-66 \pm 18	-129 \pm 26	—	—

**Fig. 9.** Free energy contour plots of the minimal units of the glycan structure. (Top left) β -D-Bac-(1 \rightarrow N)-Asn. (Top right) α -L-GalNAc-(1 \rightarrow 3)- β -D-Bac. (Bottom left) α -L-GalNAc-(1 \rightarrow 4)- α -L-GalNAc. (Bottom right) β -D-Glc-(1 \rightarrow 3)- α -L-GalNAc. Contour lines are shown at every 20 kJ/mol from 0 to 100 kJ/mol. This figure is available in black and white in print and color at *Glycobiology* online.

the network of hydrogen bonds between the catalytic residues and the acceptor Asn. In this study, we detected a dual behavior in all three PglB-Peptide simulations, 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 LLO site. It must be noted, though, that the Mg^{2+} ion is not the physiological

cation at the PglB active site—the suggested ion is Mn^{2+} . Their coordination preferences are similar (Bock et al. 1999), including the same coordination number, and similar distances (~ 0.1 Å difference). The distinction between them is that Mn^{2+} often binds in a bidentate manner, which is not common to Mg^{2+} ion (Bock et al. 1999). Still, it was detected (Lairson et al. 2008; Lizak et al. 2013) that PglB and other metalloenzymes are active in the presence of Mg^{2+} , demonstrating the importance of studying the behavior of this ion at the catalytic site. The QM calculations were employed as a refinement of the catalytic site transient states observed in the MD simulations. The presence of one water molecule at the metal coordination was confirmed by the distances and bond orders, characterizing what appears to be an octahedral geometry with six ligands that can vary the denticity between monodentate or bidentate, agreeing with the most prevalent geometry found on previous data (Bock et al. 1994; Katz et al. 1996; Venkataraman et al. 1997) and at the MESPEUS database (Hsin et al. 2008) for Mg^{2+} and Mn^{2+} . The QM final states indicated the presence of hydrogen bonds involving the Asn with Asp56 and Glu319. Also, a deviation from the planar conformation angle of the amide was detected for all frames, being frame C' (Figure 8) the highest one. In that particular case, the O δ 1 from the carboxamide of Asn is interacting with the ion Mg^{2+} , reducing the flexibility of the side chain, an important feature for the twisted amide activation hypothesis (Lizak et al. 2013). Such behavior might explain the lack of metal specificity of the enzyme. Another possibility is that the interaction between water molecules with O δ 1 and N δ 2 provides this stabilization and could break the conjugation of electrons with the nitrogen. This also would effectively raise the reactivity of the amide. All these data suggest that the coordination and catalytic mechanism presented in the crystal structure is feasible, as well as support the twisted amide activation hypothesis; however, we demonstrated at least a few novel variations in the catalytic site and in the position of the substrate.

The glycan chains were properly developed and included in two of our simulations. The glycosidic bond dihedral angles ϕ and ψ mostly agreed with the metadynamics energy plots, except for a variation that was identified in the β -D-Bac-(1 \rightarrow N)-Asn dihedral angles. The N-linkage is highly flexible, as demonstrated by the metadynamics simulation, and may present several coexisting conformational states. In our simulations of the glycosylated peptide, both inside and outside PglB, it maintained stable values near 110° for the ϕ and -150° for the ψ for that dihedral angle. We may conclude that, for the short peptide acceptor, this rigid conformation is the most prevalent one. Besides, the stable rod shape of *C. jejuni* and *C. lari* glycan chains was maintained even inside the enzyme cavities. This can raise a question: how would this substrate, linked to the lipid carrier, enter the PglB proposed glycan cavity? Yet, there is not much information regarding the LLO interaction with PglB, and how this would affect the enzyme adopted conformation. Despite recent data (Matsumoto, Shimada, Kohda, et al. 2013; Matsumoto, Shimada, Nyirenda, et al. 2013) from the Archaeal OST provided new insights on the binding of the phosphate from the LLO structure, only the obtainment of a structure with all the elements involved in the N-glycosylation, combined with studies about the implications brought by them, could clarify how this essential protein modification is orchestrated.

Materials and methods

Nomenclature and software

The IUPAC (IUPAC-IUBMB 1996) proposed recommendations for nomenclature and symbols were used. Regarding MD simulations,

the GROMACS 4.5.1 simulation suite (Pronk et al. 2013) was employed, along with the GROMOS 53A6 force field (Oostenbrink et al. 2004) for the systems without the glycan chain and the GROMOS 53A6_{GLYC} force field (Pol-Fachin et al. 2012; Pol-Fachin et al. 2014) for the systems presenting the glycan chain. The adjacent carbohydrate relative orientations are described by the dihedral angles ϕ and ψ which pertain to their glycosidic linkages. Each linkage corresponding to (1 \rightarrow X), where X is “3” or “4”, has its ϕ and ψ angles defined as shown below:

$$\phi = O5 - C1 - O1 - CX, \quad (1)$$

$$\psi = C1 - O1 - CX - C(X - 1). \quad (2)$$

For the manipulation and visualization of structures, the softwares VMD (Humphrey et al. 1996) and PyMOL (The PyMOL Molecular Graphics System) were employed.

System preparation

The crystallographic structure of the *C. lari* PglB was retrieved from the Protein Data Bank (PDB), under the code 3RCE (Lizak et al. 2011). 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 EL5, while the other is located at the PP domain, comprehending 3 residues from 605 to 607 and residing at the β -barrel-like structure. Using the web server FALC-Loop (Ko et al. 2011), 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 from the loop modeling server. Based on stereochemistry analysis by submitting the structures to PDBSUM server, which uses the PROCHECK software (Laskowski et al. 1993), we chose the model whose residues position did not block the hydrophobic groove where the LLO chain is supposed to interact. Departing from this starting structure, four systems were studied: PglB at the native unbound state (PglB-Native); PglB with a bound acceptor peptide which was already co-crystallized in the 3RCE structure (PglB-Peptide in three simulations: I, II and III); PglB with a bound glycosylated peptide portraying the heptasaccharide chain from *C. jejuni* (PglB-Glycopeptide *C. jejuni*) and PglB with a bound glycosylated peptide possessing the hexasaccharide chain from *C. lari* (PglB-Glycopeptide *C. lari*). The structure for the oligosaccharide chains was extracted from the PDB, under the code 2K33 (Slyenko et al. 2009). These glycans were rigidly docked at the LLO cavity using AutoDock 4.2 (Morris et al. 2009). We selected the minimal energy orientation and then covalently bounded with Asn.

Membrane insertion of PglB

Regarding the insertion of the protein in a membrane, the employed protocol followed the InflateGRO methodology (Kandt et al. 2007). Briefly, this consists in inserting PglB into a preequilibrated 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. The POPE membrane was previously

simulated at 310 K for 300 ns and presented an average area per lipid of $0.64 \pm 0.01 \text{ nm}^2$ (experimental data at 310 K: 0.61 nm^2) (Rappolt et al. 2003).

Molecular dynamics simulations

Following the membrane insertion steps, the rectangular box was then solvated with SPC water model (Berendsen et al. 1987) and periodic boundary conditions. Before this process, we did the common procedure of raising the van der Waals radius of the C atom from 0.15 to 0.375, aiming to avoid water molecules filling any left space between the lipids and the protein. After solvation, this parameter was returned back to its original value. Na^+ counter ions were added to neutralize the systems, when needed. The LINCS algorithm (Hess et al. 1997) 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 (Darden et al. 1993). The pressure barostat chosen was Parrinello–Rahman (Parrinello and Rahman 1981; Nosé and Klein 1983), with a 2.0 ps coupling constant, while the temperature thermostats chosen were V-rescale (NVT step) (Bussi et al. 2007) and Nosé–Hoover (NPT equilibration and production MD) (Nosé 1984; Hoover 1985), with a coupling constant of $\tau = 0.5$. Additionally, semiisotropic pressure coupling was applied, accounting for the presence of the membrane. Constant temperature of 310 K and constant pressure of 1 atm were also implemented. Steepest Descent algorithm was used in the energy minimizations performed. First, two simulations of equilibration were performed with position restraints: an NVT and an NPT of 0.1 and 0.9 ns, respectively. Subsequently, 200 ns unrestrained NPT MD simulations were performed for each of the six systems plus the glycopeptide in solution, generating the production run where data were collected for the systems analysis. The PglB–Peptide complex was evaluated in three replicates, simulations I, II and III, to increase the robustness of this conformational ensemble.

Metadynamics simulations

The metadynamics (Barducci et al. 2011) calculations were performed for the isolated disaccharides and for the Asn–Bac unit, evaluating the lowest energy of each ϕ and ψ dihedral angle. Each simulation consisted of 10 ns MD simulations, employing a height of 0.1 for the Gaussian height, and a σ of 0.5 to each of the ϕ and ψ glycosidic linkages dihedral angles. The utilized software was GROMACS 4.5.1 interfaced with the PLUMED package 1.2.2 (Bonomi et al. 2009) and the free energy surfaces were obtained by applying the *sum_bills* command.

Quantum mechanics calculations

The catalytic site residues (Asp56, Arg147, Asp154, Asp156 and Glu319), together with the Asn from the peptide, were chosen for density functional theory (DFT) (Hohenberg and Kohn 1964) calculation at B3LYP/6-31G* level. Several previous studies using B3LYP/6-31G* level were found to well reproduce geometry and energetic of reactions for different ion-containing proteins (Xu et al. 2007; Zhang et al. 2010; Abdel-Azeim et al. 2011), showing the satisfactory B3LYP accuracy for this parameters. The crystal geometry without water and three MD extracted geometries were used as start pointing for the calculations. The side chains of residues were included and the amino acids backbones were simplified with methyl groups at carbon beta. All ab initio calculations were performed using GAMESS (Schmidt et al. 1993). Briefly the geometries of the catalytic site were submitted to hydrogen geometry optimization using ab initio

quantum-mechanical computations at 3–21G level. These minimal energy conformations were submitted to DFT (B3LYP/6-31G*) level of calculations. Finally, Hessian matrix analyses were employed to unequivocally characterize the conformations obtained as true minima potential energy surfaces.

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), MCT; the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brazil; and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Acknowledgements

Research supported by the Centro Nacional de Supercomputação of the Universidade Federal do Rio Grande do Sul (CESUP/UFRGS).

Conflict of interest statement

None declared.

Abbreviations

Bac, Bacillosamine; CC, central core; DFT, density functional theory; EL, external loop; GalNAc, N-acetylgalactosamine; Glc, glucose; IS, insertion sequence; LIE, linear interaction energy; LLO, lipid-linked oligosaccharide; MD, molecular dynamics; OST, oligosaccharyltransferase; PDB, protein data bank; PME, particle mesh Ewald; PP, periplasmic; QM, quantum mechanics; RDF, radial distribution function; TM, transmembrane.

References

- Abdel-Azeim S, Li X, Chung LW, Morokuma K. 2011. Zinc–homocysteine binding in cobalamin-dependent methionine synthase and its role in the substrate activation: DFT, ONIOM, and QM/MM molecular dynamics studies. *J Comput Chem.* 32:3154–3167.
- Abu-Qarn M, Yurist-Doutsch S, Giordano A, Trauner A, Morris HR, Hitchen P, Medalia O, Dell A, Eichler J. 2007. *Halofexax volcanii* AgIB and AgID are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J Mol Biol.* 14:1224–1236.
- Barducci A, Bonomi M, Parrinello M. 2011. Metadynamics. *WIREs Comput Mol Sci.* 1:826–843.
- Berendsen HJC, Grigera JR, Straatsma TPJ. 1987. The missing term in effective pair potentials. *Phys Chem.* 91:6269–6271.
- Bock CW, Katz AK, Markham GD, Glusker JP. 1999. Manganese as a replacement for magnesium and zinc: functional comparison of the divalent ions. *J Am Chem Soc.* 121:7360–7372.
- Bock CW, Kaufman A, Glusker JP. 1994. Coordination of water to magnesium cations. *Inorg Chem.* 33:419–427.
- Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA, et al. 2009. PLUMED: a portable plugin for free energy calculations with molecular dynamics. *Comput Phys Commun.* 180:1961–1972.
- Bussi G, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. *J Chem Phys.* 126:014101–014107.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys.* 98:10089–10092.
- Eichler J. 2013. Extreme sweetness: protein glycosylation in archaea. *Nat Rev Microbiol.* 11:151–156.
- Gerber S, Lizak C, Michaud G, Bucher M, Dabre T, Aebi M, Reymond JL, Locher KP. 2013. Mechanism of bacterial oligosaccharyltransferase. *J Biol Chem.* 288:8849–8861.

- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. 1997. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 18:1463–1472.
- Hohenberg P, Kohn W. 1964. Inhomogeneous electron gas. *Phys Rev*. 136: B864.
- Hoover WG. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A*. 31:1695–1697.
- Hsin K, Sheng Y, Harding MM, Taylor P, Walkinshaw MD. 2008. MESPEUS: a database of the geometry of metal sites in proteins. *J Appl Cryst*. 41:963–968.
- Humphrey W, Dalke A, Schulten K. 1996. VMD—visual molecular dynamics. *J Mol Graph*. 14:33–38.
- 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.
- IUPAC-IUBMB. 1996. Nomenclature of carbohydrates. Recommendations. *Pure Appl Chem*. 68:1919–2008.
- Jaffee MB, Imperiali B. 2011. Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases. *Biochemistry*. 50:7557–7567.
- Kandt C, Ash WL, Tieleman DP. 2007. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*. 41:475–488.
- Katz AK, Glusker JP, Beebe SA, Bock CW. 1996. Calcium ion coordination: a comparison with that of beryllium, magnesium, and zinc. *J Am Chem Soc*. 118:5752–5763.
- Ko J, Lee D, Park H, Coutsiaris EA, Lee J, Seok C. 2011. The FALC-Loop web server for protein loop modeling. *Nucleic Acids Res*. 39:W210–W214.
- Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M. 2006. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science*. 314:1148–1150.
- Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, et al. 2006. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J*. 25:1957–1966.
- Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*. 77:521–555.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 26:283–291.
- Lizak C, Gerber S, Michaud G, Schubert M, Fan YY, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP. 2013. Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat Commun*. 4:2627.
- Lizak C, Gerber S, Numao S, Aebi M, Locher KP. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature*. 474:350–355.
- Lizak C, Gerber S, Zinne D, Michaud G, Schubert M, Chen F, Bucher M, Darbre T, Zenobi R, Reymond JL, et al. 2014. A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J Biol Chem*. 289:735–746.
- 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.
- 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.
- Matsumoto S, Shimada A, Kohda D. 2013. Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Struct Biol*. 13:11–19.
- Matsumoto S, Shimada A, Nyirenda J, Igura M, Kawano Y, Kohda D. 2013. Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci USA*. 110:17868–17873.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 16:2785–2791.
- Nosé S. 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*. 52:255–268.
- Nosé S, Klein ML. 1983. Constant pressure molecular dynamics for molecular systems. *Mol Phys*. 50:1055–1076.
- Nothaft H, Szymanski CM. 2013. Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem*. 288:6912–6920.
- Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol*. 8:765–778.
- Ollis AA, Zhang S, Fisher AC, DeLisa MP. 2014. Engineered oligosaccharyltransferase with greatly relaxed acceptor-site specificity. *Nat Chem Biol*. 10:816–824.
- Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. 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.
- Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 52:7182–7190.
- Pol-Fachin L, Rusu VH, Verli H, Lins RD. 2012. GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. *J Chem Theory Comput*. 8:4681–4690.
- Pol-Fachin L, Verli H, Lins RD. 2014. Extension and validation of the GROMOS 53A6_{GLYC} parameter set for glycoproteins. *J Comput Chem*. 35:2087–2095.
- Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 29:845–854.
- Rappolt M, Hickel A, Bringezu F, Lohner K. 2003. Mechanism of the lamellar/inverse hexagonal phase transition examined by high resolution X-ray diffraction. *Biophys J*. 84:3111–3112.
- Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, et al. 1993. General atomic and molecular electronic structure system. *J Comput Chem*. 14:1347–1363.
- Slyno V, Schubert M, Numao S, Kowarik M, Aebi M, Allain FH. 2009. NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *J Am Chem Soc*. 131:1274–1281.
- Szymanski CM, Burr DH, Guerry P. 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect Immun*. 70:2242–2244.
- Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol*. 32:1022–1030.
- The PyMOL Molecular Graphics System. Available online: <https://www.zotero.org/groups/ncdirtrd3/items/itemKey/HJ2H2FHT> (accessed on 4 August 2014).
- Varki A. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*. 3:97–130.
- Venkataraman D, Du Y, Wilson SR, Hirsch KA, Zhang P, Moore JS. 1997. A coordination geometry table of the d-block elements and their ions. *J Chem Educ*. 74:915–918.
- Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C, et al. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc Natl Acad Sci USA*. 103:7088–7093.
- Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, et al. 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298:1790–1793.
- Wiberg KB, Breneman CM. 1992. Resonance interactions in acyclic systems. 3. Formamide internal rotation revisited. Charge and energy redistribution along the C–N bond rotational pathway. *J Am Chem Soc*. 114:831–840.
- Xu D, Guo H, Cui Q. 2007. Antibiotic deactivation by a dizinc beta-lactamase: mechanistic insights from QM/MM and DFT studies. *J Am Chem Soc*. 35:10814–10822.
- Zhang C, Gao J, Chen Z, Xue Y. 2010. Molecular dynamics and density functional theory studies of substrate binding and catalysis of human brain aspartoacylase. *J Mol Graph Model*. 28:799–806.

5.2 Capítulo II

Após a obtenção da segunda estrutura de OST, desta vez uma estrutura de um organismo do domínio Archaea - a AglB, novas questões surgiram. Essa enzima apresentava unidades estruturais diferentes das obtidas nas estruturas parciais de outras arqueias. Essas unidades, a IS e a P1, ainda não tem sua função caracterizada e a única hipótese, atualmente, sugere a possibilidade de elas desempenharem um papel na termoestabilidade da enzima, considerando que diversas arqueias, como *P. furiosus*, *P. horikoshii* e a própria *A. fulgidus*, sobrevivem em temperaturas extremas. Nesse contexto, o trabalho apresentado nesse capítulo avaliou o impacto de deleções das unidades estruturais IS e P1 na manutenção da estabilidade e integridade da AfAglB-L por DM, identificando alterações conformacionais intensas no perfil das estruturas mutadas. Também foram analisados os padrões de coordenação do sitio ativo, com a intenção de verificar se as mutações afetariam as interações realizadas naquela região, o que não aconteceu.

A ideia geral é de que estudos que avaliem os efeitos da deleção dessas unidades e quais seriam suas funções nas proteínas contribuem para futuras aplicações biotecnológicas, como, por exemplo, o desenvolvimento de OSTs quiméricas que possuam unidades estruturais diferentes daquelas encontradas naturalmente.

Os resultados desse trabalho serão apresentados em formato de artigo que será submetido ao periódico *BBA General Subjects*.

Assessment of the Deletion of Structural Units in the Oligosaccharyltransferase AglB

Conrado Pedebos¹, Hugo Verli^{1*}

¹ Center for Biotechnology, Federal University of Rio Grande do Sul (UFRGS), CP 15005, Porto Alegre, RS, 91501-970, Brazil.

Email: conrado.pedebos@ufrgs.br; hverli@cbiot.ufrgs.br

* To whom correspondence should be addressed: E-mail: hverli@cbiot.ufrgs.br (H. Verli).

Tel.: +55-51-3308-6068; Fax: +55-51-3308- 7309.

Key words: Oligosaccharyltransferase, AglB, molecular dynamics, *Archaeoglobus fulgidus*

Abbreviations: Central core, CC; External loop, EL; Insertion sequence, IS; Lipid-linked oligosaccharide, LLO; molecular dynamics, MD; Oligosaccharyltransferase, OST; Peripheries 1 and 2, P1, P2; 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine, POPE; three-helices bundle, TH-Bundle; Transmembrane domain, TM.

Abstract

Background: Oligosaccharyltransferases (OSTs) are enzymes that catalyze the transfer of a glycan chain to an acceptor protein. AglB, an archaeal OST, has structural units with unknown functions. Here, we evaluate the thermostability role hypothesis by deleting units of an AglB from the hyperthermophilic species *Archaeoglobus fulgidus*.

Methods: Crystallographic structures were retrieved from the PDB databank (ID 3WAK); Membrane insertion was achieved using LAMBADA and InflateGRO2; Molecular dynamics simulations and analyses were performed employing GROMACS simulation suite with GROMOS54a7 force field.

Results: Simulations were performed for 300ns in triplicates at 356K. We observed a strong effect on the dynamics of the C-terminal domain for the mutated systems with increased fluctuations near the deleted areas. Conformational profile and stability are deeply affected, specially in the double unit deletion, modifying the enzyme behavior and binding interfaces. Coordination at the binding site was not disrupted, indicating that the mutated enzymes could retain their activity. Hotspots of variation were identified and rationalized with previous data.

Conclusions: Our data shows that the structural units can provide important stabilization interactions, as well as integrity of the wild-type enzyme at high temperatures. This work is an initial step towards the development of more evaluations involving structural units from OSTs.

General significance: The observation that structural units deletion interfere with OSTs stability and conformational preferences suggests that their role is not directly related to catalysis. Instead, they influence the OST structural integrity. These data could be exploited to interchange

structural units from different OSTs, aiming to obtain enzymes possessing novel characteristics, such as thermostability.

Introduction

N-glycosylation is a co-/posttranslational modification widespread among all domains of life. This process is involved in many actions and activities throughout cells, e. g. stability, signalling, and immune response in *Eukarya* [1]; flagella assembly, motility, protection against extreme environments, and integrity of structures in *Archaea* [2]; host adhesion, invasion, and colonization in *Bacteria* [3,4]. Despite the intrinsic differences for each species, the overall pathway share many similarities [5]. Such pathway is comprised by many enzymes that sequentially add monosaccharides to a lipid carrier, which is, then, translocated by a flippase to the exterior side of the cell (in *Archaea*), to the periplasmic space (in *Bacteria*), or to the luminal face of the ER (in *Eukarya*). In *Archaea*, the final step of the pathway is catalyzed by the oligosaccharyltransferase AglB, an enzyme that transfers the lipid-linked oligosaccharide (LLO) to an acceptor protein carrying the sequon N-X-S/T (i. e. Asn-any amino acid residue except Pro-Ser or Thr) [6].

Recently, the full structure of two distinct OSTs were elucidated: PglB from *Campylobacter lari* (CIPglB) [7], and the longest paralog AglB from *Archaeoglobus fulgidus* (AfAglB-L) (Figure 1) [8]. Before that, only partial structures for the C-terminal globular domains were obtained [9-13], demonstrating that the archaeal domain possesses the greatest variability in its structures. As seen in these previous works, the structure of an OST can be divided in two large domains: a transmembrane domain (TM) consisting of 13 α -helices, while also having external loops (EL) that display important catalytic residues [7], such as EL1, EL2, and EL5; and a C-terminal globular domain which usually consists of an α/β mixed fold. This domain can be further subdivided into structural units. Firstly, the central core (CC) unit is the

only substructure that is essential to every OST. It participates in the interface region with the

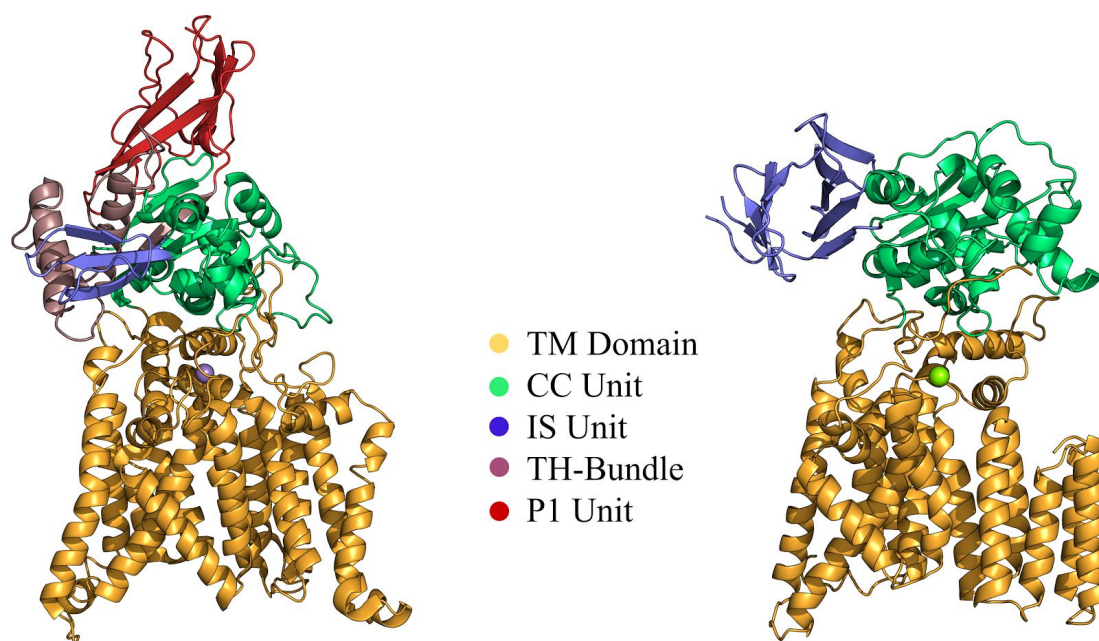


Figure 1. AfAglB-L (*left*) and CIPglB (*right*) crystallographic structures. The TM domain (orange) and the C-terminal domain structural units (CC -green, IS - blue, P1 -red) are depicted. The purple sphere corresponds to Zn^{2+} ion, while the green sphere corresponds to Mn^{2+} ion. The TH-Bundle is originally part of CC unit, although it occupies the same place of the IS unit from CIPglB and interacts with the IS unit from AfAglB.

TM domain, assembling the enzyme two binding sites. Accordingly, this unit holds the WWDYGY (WWDXGX) conserved motif [9,14], responsible for the recognition of the S/T (+2) of the N-glycosylation sequon, and the DK motif [7,9], which supports the maintenance of this residue in a tight binding. The second unit is the Insertion Sequence (IS), which receives this name because it seems to be inserted in the CC unit area [13]. Generally, it displays an

organization of β -strands. For example, CIPglB and *P. furiosus* AglB (PfAglB) possess a distorted β -barrel fold, while AfAglB-L presents a reduced unit without the β -barrel organization. As a substituent for the barrel structure, a three-helices bundle (TH-Bundle) (originally classified as being part of the CC unit) [13] is located at the same place and seems to play that role, a unique feature in the OSTs structures found until now, including the partial structures. The two final structural units are the Peripheries 1 and 2, only identified in Archaeal structures. These units are β -sheet rich, with slightly different organizations and they are distributed around the CC unit, on the opposite direction of the IS unit. Archaeal OSTs can display either only P1 or both units at the same time, but no structure carrying only P2 was found yet.

Despite all of this information, until today only one experimental study assessed structural units properties [10]. In this previous work by Matsumoto *et al.*, a mutant PfAglB with a deleted IS unit (PfAglB Δ IS) displayed virtually the same activity than the wild type enzyme, indicating that this unit is dispensable for catalysis. However, the assays were performed with detergent-solubilized PfAglB and a short peptide as an acceptor, caveats that could interfere with proper evaluation of the enzyme function [10]. Two hypotheses emerged for the functions of the IS unit: structural stability when the protein is embedded in a membrane, and support for the regulation or efficiency of the enzyme's activity when in the presence of larger polypeptide chains and when coupled to protein translation processes [10]. As for P1 and P2, the only current hypothesis [13] is that they could contribute for the thermostabilization of the enzyme under the extreme conditions in which archaeal species survive, since the extra units (and specifically P1 and P2) are only found in this domain of life. Another mentioned possibility is that these units could be equivalent to the subunits that constitute the OST multimeric complex from eukaryotic

species (5-8 subunits), however the limited sequence similarity between them hinders further predictions [10].

Investigations of OSTs mechanisms could lead to the production of tailor-made glycoproteins. In fact, recent data have demonstrated modifications which enhanced OSTs catalysis, such as mutations of residues to increase the recognition and binding of acceptors [15-17], change of the catalytic ion to increase the rate of product formation [8,18,19], as well as the production of different sizes of lipid carrier and interspecies glycan chains [20,21]. However, there is still a lack of studies evaluating the role of IS, P1, and P2 structural units, which could bring new insights into the enzyme functionality. Therefore, in this work, we employed molecular dynamics (MD) simulations to observe the structural effects caused by full deletions of AfAglB-L extra structural units. Specifically, we aimed to: I) Evaluate the mutants stability after the deletion of P1, and P1 and IS units together. II) Verify the behavior of the catalytic site in order to analyze if the requirements for catalysis are maintained; III) Identify important interactions between structural units that govern the protein stability under hyperthermophilic conditions, allowing inferences regarding these units transferability between distinct OSTs. Our work seeks to contribute for the comprehension of the structural basis that governs OSTs dynamical behavior, providing insights for potential biotechnological applications of these enzymes.

Materials and methods

Systems Preparation

For this study, the structure from RCSB Protein Data Bank under the PDB-ID: 3WAK [8]

was retrieved, since it corresponds to a full apo structure containing EL5. Mutants were obtained by deleting the P1 region (residues 777-868), and the IS region plus the TH-Bundle (residues 638-776), including a short region from the CC unit that connects IS to P1. Three systems were studied in this work: i) Wild type AfAglB-L in its apo state (AfAglB-L-WT); ii) AfAglB-L with the deletion of P1 unit (AfAglB-L- Δ P1); iii) AfAglB-L with deletion of IS and P1 (AfAglB-L- Δ ISP1). In this work, we established that the Δ ISP1 included the TH-Bundle from the CC unit. The reason for that is because it makes close contacts with the IS unit and forms a structure separated from the enzyme core, resembling the common β -barrel IS seen in other OSTs. To better describe the environment where these proteins are found, they were embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) membrane with 512 phospholipids. The methodology utilized for protein insertion was based on the tools LAMBADA and InflateGRO2 [22].

Molecular Dynamics Simulations

The triclinic simulation boxes received SPC water solvation [23], in the presence of periodic boundary conditions. In the solvation step, Van der Waals radii of C atoms were raised from 0.150 to 0.375 to avoid water molecules filling spaces between phospholipids and the protein. This parameter was returned to the default value after this step. PME [24] was chosen for the electrostatic treatment of the system. The LINCS method [25] was applied to constrain covalent bonds, allowing an integration step of 2 fs. GROMACS simulation suite [26] with GROMOS54A7 force field [27] was used for the molecular dynamics (MD) simulations. POPE membranes were described by GROMOS53A6 parameters (compatible with GROMOS54A7)

following a previous report [28]. Cl^- counter ions were added to neutralize the residual charge of the systems, when needed.

Firstly, energy minimizations were performed using the Steepest Descent algorithm. Then, MD simulations were divided in two phases: an equilibration phase and a production phase. For the equilibration phase, systems were simulated for 1 ns in a canonical ensemble (NVT) and for 19 ns in an isothermal-isobaric ensemble (NPT). Parrinello-Rahman barostat [29,30] was applied with a coupling constant of 2.0 ps. As thermostats, V-rescale (NVT step) [31] and Nosé-Hoover (NPT step - equilibration and production phases) [32,33] with coupling constants of $\sigma = 0.1$ and 0.8, respectively. Also, semiisotropic pressure was applied due to the presence of the membrane. A constant temperature of 356 K was implemented, since this is the optimal growth temperature for *A. fulgidus* [34]. During equilibration phase, the protein was subjected to position restraints with a force constant of $1000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$. Subsequently, in the production phase, unrestrained MD simulations were performed for 300 ns for each system. Simulations were performed in triplicates (3 x 300 ns), starting with different velocities, aiming to enhance the sampling robustness of the conformational ensemble. Data was collected during the production phase and analyzed employing GROMACS suite tools. VMD [35] and PyMOL [36] were used for trajectories visualization and for manipulation of structures.

Results

Global Structural Stability

We began by performing a global evaluation of the enzymes observing the deviation from the initial structures. The root mean square deviation (RMSD) (Figure 2A) plot for the whole

structure indicates that, despite having an average difference of 0.1 nm between AfAglB-L-WT and AfAglB-L- Δ P1, the large standard deviation prevents us from asserting a significant difference between both systems. In contrast, the AfAglB-L- Δ ISP1 mutated system appears to suffer from a dramatic increase on its average RMSD value after 200 ns of simulation time, reaching an average of \sim 1.1 nm near the end. In this case, in spite of its high standard deviations, the values are significantly different when compared to the other two systems. Analyzing the root mean square fluctuation (RSMF) (Figure 3) of the backbone atoms as an average per residue of the concatenated trajectories, we observed that the TM domain (residues 1 to 500) maintain a generally stable behavior in its helices for all systems, with some instability for AfAglB-L- Δ ISP1. The intracellular loops display higher fluctuations for all systems; EL5 also shown its expected flexibility. Particularly, the intracellular loop region (residues 454 to 464) connecting the last two helices of this domain demonstrated an increase of 0.3 nm in AfAglB-L- Δ ISP1 when compared to the other systems. Analyzing the C-terminal domain, we perceived a deeper influence from the structural units deletion. In the mutated systems, there is a similar pattern of higher fluctuation when compared to AfAglB-L-WT, augmenting the flexibility in specific regions of the CC unit, especially pronounced in AfAglB-L- Δ ISP1: a helix (residues 517-522) and a loop (residues 533-543) (Figure 3) that were previously at the vicinity of P1 unit; two nearby loops (residues 569-571 / residues 576-582) in the middle of CC; another helix, in AfAglB-L- Δ ISP1, which was next to both deleted units; the kinked helix, which carries the DK motif; the short helix-loop final portion of CC unit. IS unit seems to be naturally flexible, since it demonstrated some flexibility also in the AfAglB-L-WT system. However, the deletion of P1 caused an enhancement of this feature, more specifically at the TH-Bundle, which interacts with

the β -hairpin structure, substituting the distorted β -barrel seen in CIPglB. To further evaluate the extent of these instabilities, we analyzed the secondary structure content of each system (Figure 2B), considering only the remaining residues in common between all systems. All proteins had a decrease on their content of at least 20 residues. Both mutations appears to have a higher decrease trend on the secondary structure content, specially in AfAglB-L- Δ ISP1 exhibiting the most pronounced, and significant, diminishment (around 35 residues).

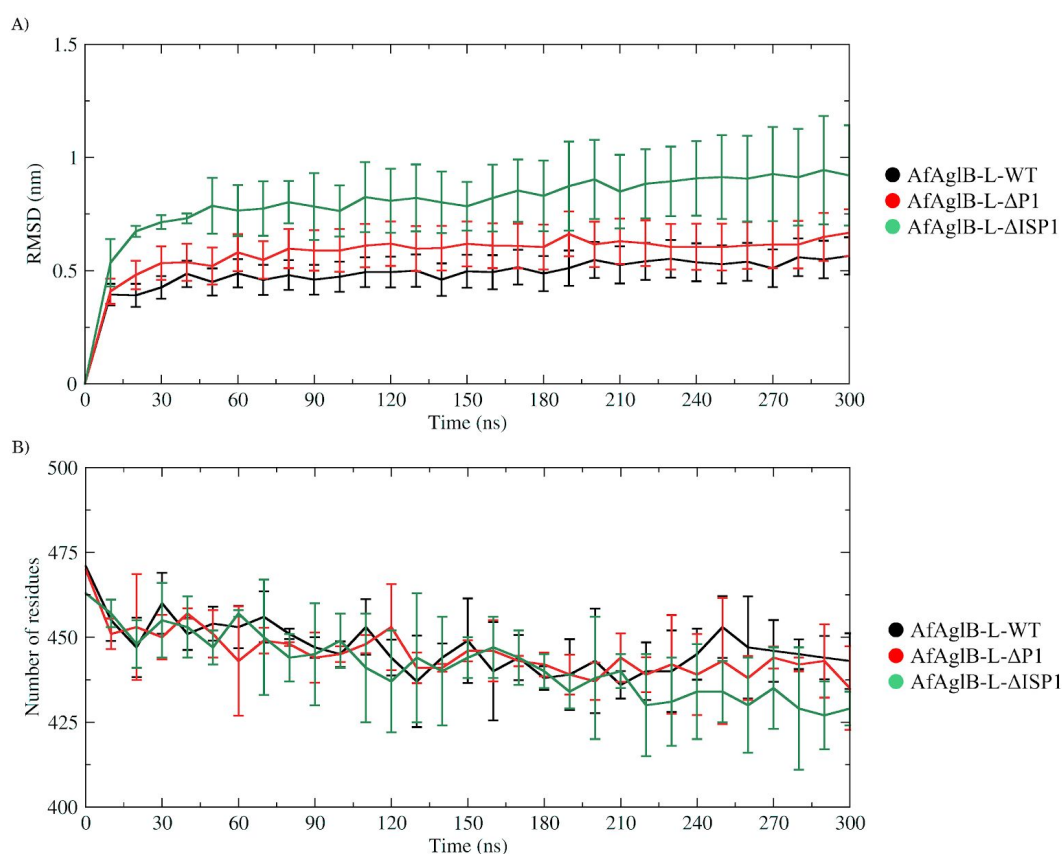


Figure 2. RMSD and Secondary Structure content during simulations. A) RMSD relative to the crystal structure for each system at every 10ns with standard deviations. B) Secondary structure content for each system at every 10ns with standard deviations.

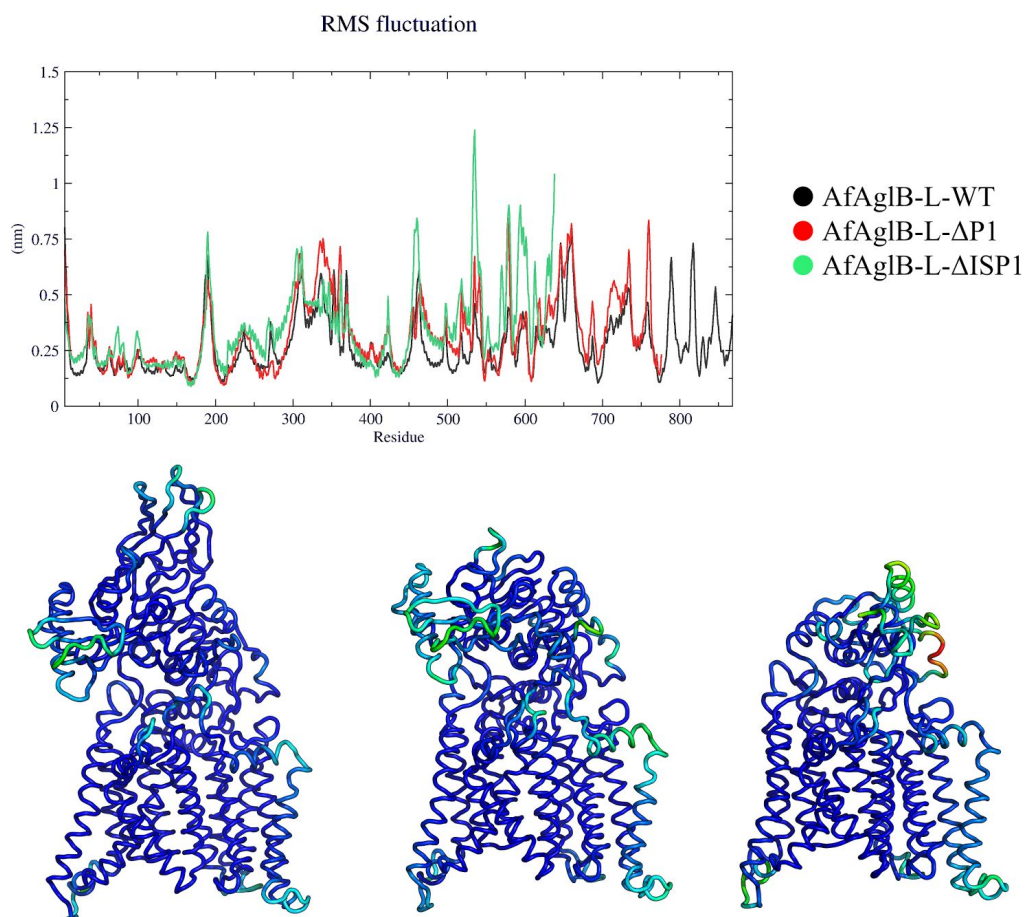


Figure 3. Individual backbone RMS fluctuation for each residue of every system. *Top.* RMS fluctuation of the backbone structure as an average per residue of concatenated replicates for each system. *Bottom.* Residues colored by the intensity of the fluctuation in a qualitative comparison with a heat map color code (blue = low fluctuation; red = high fluctuation).

Structural Units Deletion Affects the Conformational Behavior of AfAglB

To better comprehend how the mutations impact on the sampling of the conformational landscape of AfAglB, we employed principal component analysis (PCA) on the MD simulations

data. By plotting the two-dimensional subspace defined by the top two principal components (PCs) (Figure 4), we were able to evaluate the structural relaxation of the proteins individually. For the WT system, a similar behavior occurred for all replicates considering the whole protein structure, as observed by plotting the first and the second eigenvectors projections: an initial fluctuation, jumping among minimum energy basins, as well as an apparent accommodation by the end of each simulation. Replicates 1 and 2 sample mostly one specific low-energy state, while replicate 3 takes some time jumping among different states before finding a more stable minimum. Despite that, the replicates explored distinct areas from the subspace, indicating that our simulations did not reach convergence in the time performed. As for the mutated systems, the replicates behavior are identical: a large drift through the conformational space, with almost no detectable minimum energy basins, as well as no overlap between systems, indicating that these systems are even farther from convergence than the apo state simulations.

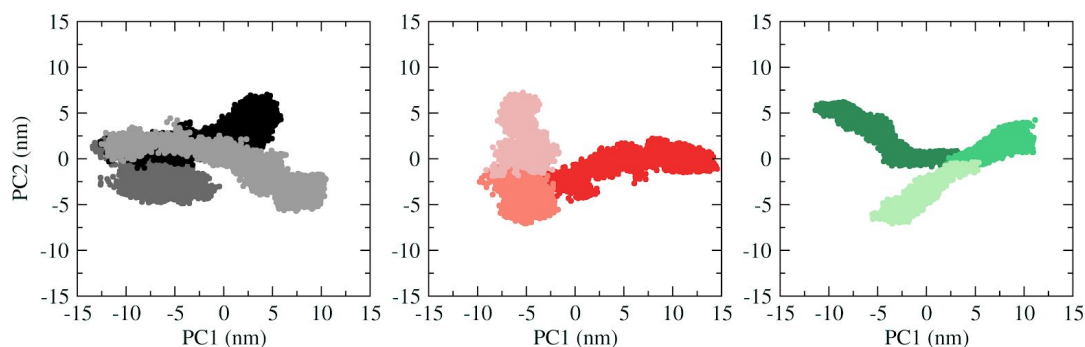


Figure 4. Top two PCs projections extracted with PCA technique for the individual systems with concatenated trajectories. Individual replicates are represented by a color code of shades of black (AfAglB-L-WT), red (AfAglB-L- Δ P1), and green (AfAglB-L- Δ ISP1).

To better characterize the influence of the deletions on the protein structure, we employed

the same analysis for the isolated TM domain, and for the C-terminal domain structural units (CC, IS, and P1) of AfAglB-L (Figure 5). Firstly, all trajectories are concatenated into one, so that the results may be compared for a common subspace. Similarly to what we saw for the RMSD and RMSF plots, AfAglB-L-WT and AfAglB-L- Δ P1 demonstrated a comparable behavior, including an apparent level of overlap between some of their replicates, specifically in the TM domain and in the CC unit. In addition, these units movements are more contained at specific portions of the subspace spanned by the two PCs analyzed, a behavior which suggests that there is less conformational variation in this domain. Contrastingly, AfAglB-L- Δ ISP1 is largely affected by the dual deletion, demonstrating the same drift behavior observed for the global structure, clearly not converging during the simulation time in most of its replicates. In fact, the structural adjustment suffered by AfAglB-L- Δ ISP1 is so intense that subtle conformational variations from the other two systems are concealed. The CC unit repeated the same pattern seen for the TM domain: AfAglB-L-WT and AfAglB-L- Δ P1 displayed a similar shape on the projections, while AfAglB-L- Δ ISP1 maintained its dominance in the analysis with a high drifting pattern, even larger than for the TM domain, confirming that this mutation affected more the C-terminal area. Subsequently, when we analyzed the IS unit, the deletion in AfAglB-L- Δ P1 demonstrated a considerable amount of influence on this unit behavior. Replicate 2 mainly sampled the same regions as the replicate 2 of the WT system, while replicates 1 and 3 from AfAglB-L- Δ P1 drifted apart from the others, accounting for most of the variance constituting both PCs. AfAglB-L-WT replicate 3 samples the same area as replicate 1 from the same system, while replicate 2 generates a new basin. Finally, P1 unit (from Af-AglB-L-WT) shown a less diffuse behavior, possessing one well-defined minimum energy state where the

three replicates occupied for some time during the simulations, and two other states that were occupied separately by replicates 2 and 3, while replicate 1 stayed trapped on the first minimum.

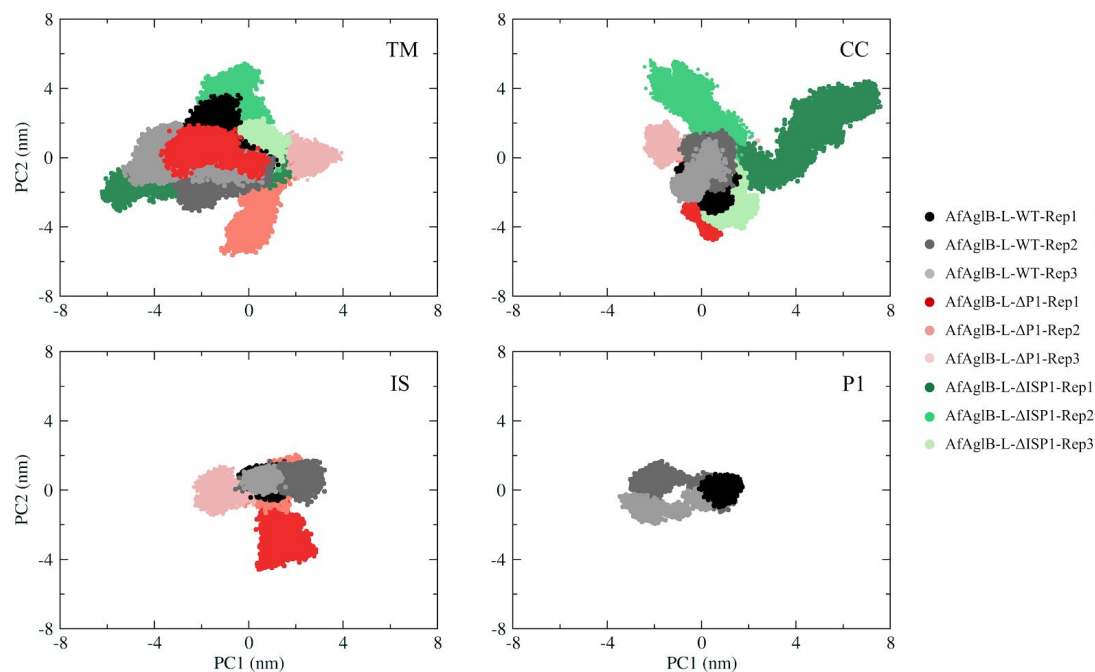


Figure 5. Top two PCs projections for all systems with concatenated trajectories extracted with PCA technique. Individual replicates are represented by a color code of shades of black (AfAglB-L-WT), red (AfAglB-L- Δ P1), and green (AfAglB-L- Δ ISP1).

Binding Site Coordination and Interface Stability

An important feature of OSTs, as metalloenzymes, is the maintenance of an organized network of interactions between the catalytic residues, as well as the metal ion coordination geometry. Consequently, we evaluated these interactions during the MD simulations by calculating the distances concerning the atoms from the side-chain of essential residues (Table 1). From that, it was possible for us to observe that all systems maintained a similar coordination geometry to the one found on the crystallographic structure (PDB ID=3WAK), although with a

couple of differences: His163, which showed an increase on its distance relative to Zn^{2+} , and the presence of more water molecules surrounding the metal ion (3:1), similarly to the coordination proposed on crystal form 1 in the same work by Matsumoto *et al.* (PDB ID=3WAJ) [8]. Glu360, a presumed critical residue for the catalysis mechanism (equivalent to Glu319 in ClPglB), did not interact with the metal ion, thus not forming the complete active site, as detected in the crystallographic structure. The metal ion remained in a tight binding at the enzyme core, with very rigid distances between the metal ligands (max. standard deviation of 0.4 Å), and one of the water molecules bridging an interaction involving His163 and Zn^{2+} . We identified fluctuating coordination numbers, between 6 and 7, depending on the denticity of the carboxylates from Asp and the number of water molecules involved (Figure 6), exhibiting either a distorted octahedral shape or a face-capped octahedron.

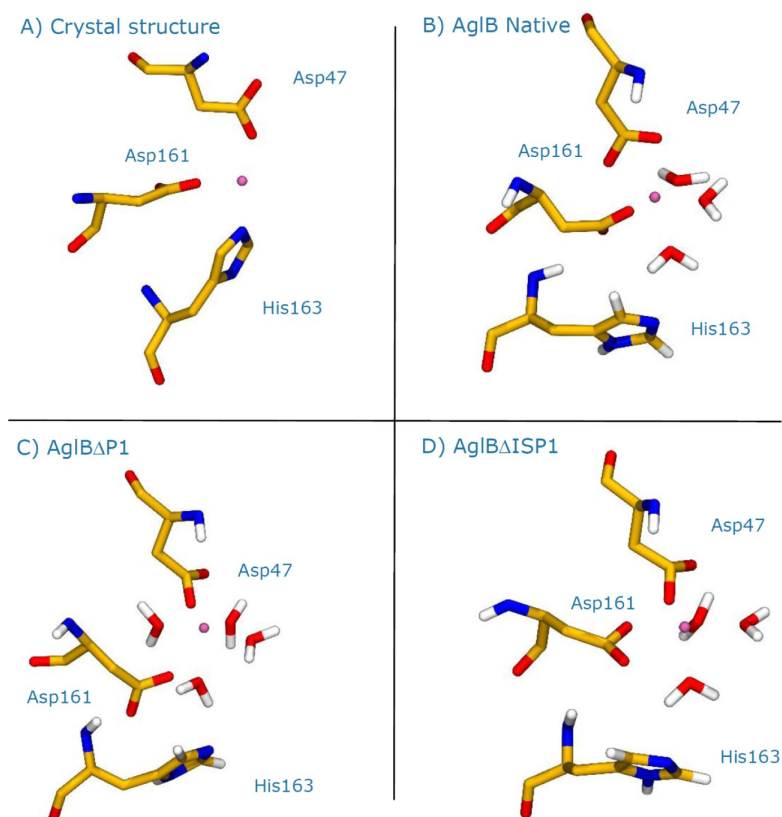


Figure 6. Snapshots of representative coordination geometries for the crystal structure and each system studied. Residues from the catalytic site are represented in yellow licorice, while the rest of the protein, the membrane, and other water molecules are hidden for better visualization.

Discussion

The deletion of structural units from the C-terminal domain of AfAglB-L clearly impacts on many aspects of the enzyme. The TM domain appears to be the least affected one, as we expected, since it is far from the mutations area, besides being embedded in the lipid environment which provides stabilizing interactions. As seen in the RMSD and RMSF plots, the CC unit of AfAglB-L- Δ ISP1 was deeply affected by the deletions, showing larger fluctuations throughout

this unit, including modifying the behavior of the kinked helix that carries the DK motif. As tested by previous reports [11,37], this modification could negatively influence sequon recognition by the OST, rendering a poor protein binding, and, ultimately, reduce the enzyme activity. In a lesser extent, AfAglB-L- Δ P1 also suffers from the effects observed in the previous mutant. The destabilization pattern is very similar in the CC unit, repeating the same instability at the DK motif. The increased flexibility of the IS unit may pose as a drawback because not only it modifies the binding site interface with the acceptor protein, but it also propagates to the CC unit, further destabilizing it in many areas, including the DK motif and the area near the LLO cavity. The presence of P1 in AfAglB-L-WT stabilizes the IS unit by the interface between the two units, interacting with helix α C of the TH-Bundle from that unit and the loop connecting it to the CC unit, as well as the loop from the opposite side (above the LLO cavity).

PCA technique is classified as a valuable tool to assess conformational sampling, and convergence of simulations [38,39,40]. In our simulations, by using PCA, we identified important aspects involving the conformational ensembles of the studied systems: 1) In spite of none of the systems reaching convergence during our simulations, AfAglB-L-WT was the most dynamically stable system, since it was able to achieve minimum energy basins, sample this areas, and has more overlap between its replicates, as well as a smaller drift along the PCs. The mutations affected the behavior of AfAglB-L, leading to huge drifts along the conformational space with few clearly defined basins being explored. 2) When evaluated in the same subspace, it is evident that the double unit deletion has a heavier impact on protein structure than the single deletion. In this case, none of the domains from AfAglB-L- Δ ISP1 demonstrated a stable behavior, keeping the same projections pattern observed for the whole protein. AfAglB-L-WT

and AfAglB-L- Δ P1, contrarily, exhibited some superimposition, which is probably due to the dominance of AfAglB-L- Δ ISP1 largely fluctuating motions. Individual analyses for AfAglB-L-WT and AfAglB-L- Δ P1 systems confirmed that the overlap between their replicates is low, as well as highlighted the instabilities in the conformational profile of AfAglB-L- Δ P1 when compared to WT. Therefore, we conclude that the mutated systems are mostly unable to keep a similar conformational behavior to AfAglB-L-WT, possibly interfering with the enzyme action, since the specific conformations that allow the binding of substrates may not be available on these systems.

Coordination of the metal ion at the binding site was very stable, indicating that, despite all the disturbance in the protein structure caused by the deletions, the enzyme could still be catalytic competent, as long as the two substrates manage to bind in the cavities. A puzzling aspect of the coordination at the catalytic site involves His163. This residue is coordinating the metal ion in the crystal structure, displaying a distance of 2 Å from the nitrogen atom Nε2 to Zn²⁺. It is part of the DXD motif (DXH in AfAglB-L) of glycosyltransferases, equivalent to Asp156 in CIPglB OST, which has not been identified as one of the coordinating residues in its crystallographic structure [7]. These structures were obtained in different states, the apo (AfAglB-L) and the cocrystallized with the peptide acceptor (CIPglB), therefore, one could assume that Asp156 and His163 may have a coordinating role only when the protein is in its apo state. However, both in this work and in a previous report from CIPglB simulations [41], we did not observe this coordination participation during our simulations. Instead, a bidentate coordination by Asp47 (Asp56 in CIPglB) and water molecules mediated this interaction, maintaining the octahedral coordination geometry. Differently, a recent work [42] also involving

MD simulations of PglB, observed a coordination participation from Asp156 in all states studied. This way, the role of these residues is not clear yet, despite being widespread in this enzyme family [43], as well as being fundamental to the enzyme activity [44]. One possibility is that His163 (and Asp156 in CIPglB) serves as a “backup” coordination ligand, preventing the metal ion from escaping the catalytic site when e.g. the protein substrate is absent, and/or bridging water molecules interactions. Another possibility would be that this residue is involved in catalysis by supporting the binding or the stabilization of the LLO substrate, since the monosaccharide at the reducing-end of the glycan is a glucose with three sulfates in positions C2, C3 and C6 and could be accommodated in a pocket consisting of the conserved residues His81, His162, His163, and Trp215. Of course, we have to consider that MD simulations do not handle electronic properties directly, which could be a caveat that prevented us from observing a correct description of the His163 interactions. Glu360 never reached the metal ion during any of the simulations, hence, the binding site was never fully complete. This strengthens the hypothesis that acceptor protein binding is important for the formation of the complete binding site [7,8].

The data obtained for the catalytic site coordination corroborates previous data from PfAglB, where the catalytic activity of PfAglB- Δ IS was not affected by the deletion [10]. Contrastingly, the global motions and the binding interface, as seen by the PCs motions on its extreme projections (Figure S1), disagree with these observations. In AfAglB-L-WT, motions are dominated by EL5 from the TM unit, and by a rearrangement of the IS unit, similarly to motions identified in CIPglB dynamics [41], reinforcing the observations that the IS unit is intrinsically flexible. In AfAglB-L- Δ P1, there is a twisting motion between TM and PP domains, whereas, in AfAglB-L- Δ ISP1, we identified huge collective motions, specially on the PP domain,

disorganizing the protein binding site. Comparing the PP domain of the wild-type structures from both species (Figure 7), we identified possible explanations for these discrepancies: the IS unit from PfAglB is a stable β -barrel structure, and the deletion (residues 603-678) allows residues 602 and 679 to be easily connected, thus causing less perturbations in the enzyme core than our deletion in AfAglB-L-WT (including the TH-Bundle and one conserved β -sheet at the enzyme core). Indeed, the deletion in PfAglB permits a structural organization of the CC unit that resemble the other two paralogs from *A. fulgidus*, AfAglB-S1 and AfAglB-S2, which do not portray IS units. Furthermore, PfAglB possesses two periphery units (P1 and P2), one that occupies the same place as P1 in AfAglB-L-WT, and other that encircles the area above the LLO cavity. As seen in our RMSF plots (Figure 3), this area has many unstable regions with fluctuations that are largely increased as a side-effect of the deletions. Flexibility of the same nature, at the same regions, was also seen in ClpG [41] MD simulations, confirming that these are sensitive areas. Possibly, the presence of the β -strand rich P2 unit provides additional stabilization for PfAglB, protecting CC unit from unfolding, as well as allowing the deletion of IS to be tolerable. Perhaps, if AfAglB-L-WT possessed both P1 and P2, it could also tolerate the deletion of IS, as the single deletion of P1 was less aggressive to the structural organization of the enzyme core than the double deletion. Besides, the large-amplitude motions observed after both deletions in AfAglB-L would be contained by the presence of P2, promoting diminished fluctuations at the PP domain.

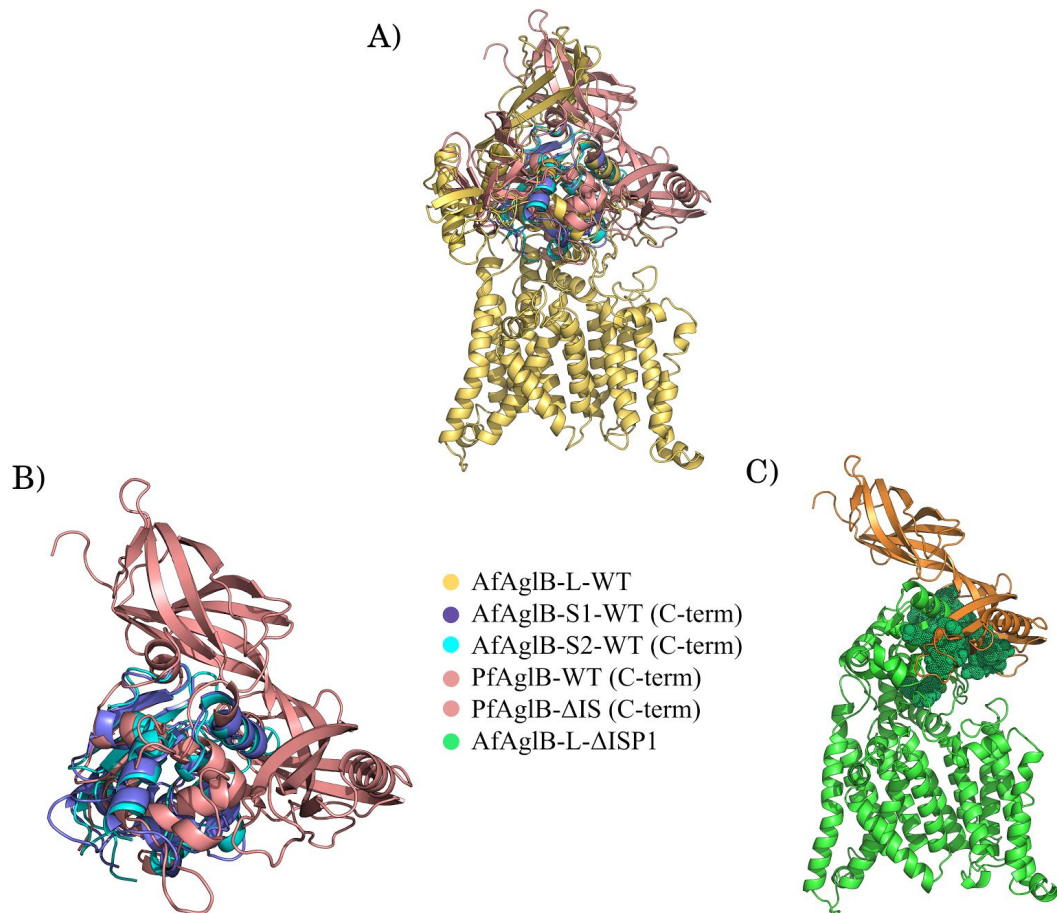


Figure 7. Structural comparison between AfAglB-L, AfAglB-S1 and S2, and PfAglB. A) AfAglB-L with all other C-terminal domains aligned to its corresponding domain. B) PfAglB with the deletion of IS unit folds in a way that resembles the other two paralogs from *A. fulgidus*. C) Alignment of the C-terminal domain of PfAglB-WT to AfAglB- Δ ISP1 left the P1 and P2 units (orange) aligned near a sensible area (green spheres) of the mutated system (AfAglB- Δ ISP1), possibly protecting that region in a hypothetical chimeric AfAglB-L.

Conclusions

Structural studies of OSTs have been in constant advancement, enabling a better understanding of the molecular basis involving N-glycosylation modifications, which, in turn, may provide new tools for the glycoengineering of biomolecules. Also, many experiments have been performed regarding these enzymes catalytic mechanism, usually evaluating residues from CC unit and TM domain, as well as studies designed to enhance their activities. However, to this date, only one previous work has assessed the role of other structural units from OSTs [10]. In this work, we try to contribute for the structural comprehension of OSTs and increase the knowledge involving structural units dynamics and function. By employing MD simulations, we could verify that both P1 and IS units provide fundamental stabilizing interactions that permits both structural integrity of AfAglB-L in extreme environments, such as the high temperatures faced by *A. fulgidus*, along with the preservation of the enzyme conformational profile and binding interface, important features for its proper functioning. Combining our findings with previous data from literature, we identified that there might be some interchangeability between structural units. For example, one could transfer a P1 unit from AfAglB or P1 and P2 from PfAglB to another OST, aiming to achieve a new class of engineered chimeric thermostable OSTs. PglB from the *Campylobacter* or *Desulfovibrio* genera would be good candidates for this approaches, since they are found in mesophilic species and have a well-known promiscuity [17]. We hope that this work support the development of more assessments regarding OSTs structural units.

Acknowledgments

Research was supported by the Centro Nacional de Supercomputação of the Universidade Federal do Rio Grande do Sul (CESUP/UFRGS), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior), and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul).

Conflict of interest statement

None declared.

References

- [1] A. Varki, Biological roles of oligosaccharides: all of the theories are correct, *Glycobiology*. 3 (1993) 97–130.
- [2] J. Eichler, M. Koomey, Sweet New Roles for Protein Glycosylation in Prokaryotes, *Trends Microbiol.* S0966-842 (2017) 30053-30067.
- [3] C.M. Szymanski, D.H. Burr, P. Guerry, *Campylobacter* protein glycosylation affects host cell interactions, *Infect Immun.* 70 (2002) 2242–2244.
- [4] H. Nothaft, C.M. Szymanski, Bacterial protein N-glycosylation: new perspectives and applications, *J Biol Chem.* 288 (2013) 6912–6920.
- [5] F. Schwarz, M. Aebi, Mechanisms and principles of N-linked protein glycosylation, *Curr Opin Struct Biol.* 21 (2011) 576-582.
- [6] B. Chaban, S. Voisin, J. Kelly, S. M. Logan, K. F. Jarrell, Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked

- glycosylation pathways in Archaea, *Mol Microbiol.* 61 (2006) 259-268.
- [7] C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, X-ray structure of a bacterial oligosaccharyltransferase, *Nature.* 474 (2011) 350-355.
- [8] S. Matsumoto, A. Shimada, J. Nyirenda, M. Igura, Y. Kawano, D. Kohda, Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation, *Proc Natl Acad Sci U S A.* 110 (2013) 17868-17873.
- [9] N. Maita, J. Nyirenda, M. Igura, J. Kamishikiryo, D. Kohda, Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases, *J Biol Chem.* 285 (2010) 4941-4950.
- [10] S. Matsumoto, M. Igura, J. Nyirenda, M. Matsumoto, S. Yuzawa, N. Noda, F. Inagaki, D. Kohda, Crystal structure of the C-terminal globular domain of oligosaccharyltransferase from *Archaeoglobus fulgidus* at 1.75 Å resolution, *Biochemistry.* 51 (2012) 4157-4166.
- [11] M. Igura, N. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka, D. Kohda, Structure-guided identification of a new catalytic motif of oligosaccharyltransferase, *EMBO J.* 27 (2008) 234-243.
- [12] J. Nyirenda, S. Matsumoto, T. Saitoh, N. Maita, N.N. Noda, F. Inagaki, D. Kohda, Crystallographic and NMR evidence for flexibility in oligosaccharyltransferases and its catalytic significance, *Structure.* 21 (2013) 32-41.
- [13] S. Matsumoto, A Shimada, D. Kohda, Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases, *BMC Struct Biol.* 13 (2013) 11-20.
- [14] Q. Yan, W.J. Lennarz, Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process, *J Biol Chem.* 277 (2002) 47692-47700.
- [15] A.A. Ollis, S. Zhang, A.C. Fisher, M.P. DeLisa, Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity, *Nat Chem Biol.* 10 (2014) 816-822.

- [16] J. Ihssen, J. Haas, M. Kowarik, L. Wiesli, M. Wacker, T. Schwede, L. Thöny-Meyer, Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering, *Open Biol.* 5 (2015) 140227.
- [17] A.A. Ollis, Y. Chai, A. Natarajan, E. Perregaux, T. Jaroentomeechai, C. Guarino, J. Smith, S. Zhang, M.P. DeLisa, Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences, *Sci Rep.* 20 (2015) 15237.
- [18] S. Gerber, C. Lizak, G. Michaud, M. Bucher, T. Darbre, M. Aebi, J.L. Reymond, K.P. Locher, Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis, *J Biol Chem.* 288 (2013) 8849-8861.
- [19] C. Lizak, S. Gerber, G. Michaud, M. Schubert, Y.Y. Fan, M. Bucher, T. Darbre, M. Aebi, J.L. Reymond, K.P. Locher, Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation, *Nat Commun.* 4 (2013) 2627.
- [20] F. Liu, B. Vijayakrishnan, A. Faridmoayer, T.A Taylor, T.B. Parsons, G.J. Bernardes, M. Kowarik, B.G. Davis, Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation, *J Am Chem Soc.* 136 (2014) 566-569.
- [21] A. Srichaisupakit, T. Ohashi, R. Misaki, K. Fujiyama, Production of initial-stage eukaryotic N-glycan and its protein glycosylation in *Escherichia coli*, *J Biosci Bioeng.* 119 (2015) 399-405.
- [22] T.H. Schmidt, C. Kandt, LAMBADA and InflateGRO2: efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations, *J Chem Inf Model.* 52 (2012) 2657-2666.
- [23] H.J.C. Berendsen, J.R. Grigera, T.P.J. Straatsma, The missing term in effective pair potentials, *Phys Chem.* 91 (1987), 6269–6271.
- [24] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N·log (N) method

- for Ewald sums in large systems, *J Chem Phys.* 98 (1993) 10089–10092.
- [25] B. Hess, P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation, *J Chem Theory Comput.* 4 (2008) 116-122.
- [26] M.J Abraham, T. Murtolad, R. Schulzb, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX.* 1-2 (2015) 19-25.
- [27] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, *Eur Biophys J.* 40 (2011) 843-856.
- [28] T.J. Piggot, D.A. Holdbrook, S. Khalid, Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes, *J Phys Chem B.* 115 (2011) 13381-13388.
- [29] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J Appl Phys.* 52 (1981) 7182–7190.
- [30] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol Phys.* 50 (1983) 1055–1076.
- [31] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J Chem Phys.* 126 (2007) 014101–014107.
- [32] S. Nosé S A molecular dynamics method for simulations in the canonical ensemble, *Mol Phys.* 52 (1984) 255–268.
- [33] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions,

Phys Rev A. 31 (1985) 1695–1697.

[34] H-P. Klenk, R.A. Clayton, J-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, S. Peterson, C.I. Reich, L.K. McNeil, J.H. Badger, A. Glodek, L. Zhou, R. Overbeek, J.D. Gocayne, J.F. Weidman, L. McDonald, T. Utterback, M.D. Cotton, T. Spriggs, P. Artiach, B.P. Kaine, S.M. Sykes, P.W. Sadow, K.P. D'Andrea, C. Bowman, C. Fujii, S. A. Garland, T.M. Mason, G.J. Olsen, C.M. Fraser, H.O. Smith, C.R. Woese, J.C. Venter, The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, Nature. 394 (1998) 364-370.

[35] W. Humphrey, A. Dalke, K. Schulten, VMD—visual molecular dynamics, J Mol Graph. 14 (1996) 33–38.

[36] The PyMOL Molecular Graphics System. Available online:

<https://www.zotero.org/groups/ncdirtrd3/items/itemKey/HJ2H2FHT> (accessed on July 11, 2017).

[37] K. Hese, C. Otto, F.H. Routier, L. Lehle, The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*, Glycobiology. 19 (2009) 160-171.

[38] H.W. Ng, C.A. Laughton, S.W. Doughty, Molecular Dynamics Simulations of the Adenosine A2a Receptor: Structural Stability, Sampling and Convergence, J Chem Inf Model. 53 (2013) 1168-78.

[39] A. Amadei; M.A. Ceruso, A. Di Nola, On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations, Proteins. 36 (1999) 419-424.

[40] B. Hess, Convergence of sampling in protein simulations, Phys Rev E Stat Nonlin Soft Matter Phys. 65 (2002) 031910-031920.

- [41] C. Pedebos, P.R. Arantes, G.M. Giesel, H. Verli, *In silico* Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination, *Glycobiology*. 25 (2015) 1183–1195.
- [42] H.S. Lee, W. Im, Transmembrane Motions of PglB Induced by LLO are Coupled with EL5 Loop Conformational Changes Necessary for OST Activity, *Glycobiology*. 2017, doi: 10.1093/glycob/cwx052
- [43] C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča, A. Imberty, Structures and mechanisms of glycosyltransferases, *Glycobiology*. 16 (2006) 29R-37R.
- [44] J. Li, D.M. Rancour, M.L. Allende, C. A. Worth, D.S. Darling, J.B. Gilbert, A.K. Menon, W.W. Young Jr, The DXD motif is required for GM2 synthase activity but is not critical for nucleotide binding, *Glycobiology*. 11 (2001) 217-229.

Table 1. Distances from catalytic residues to Zn²⁺ ion during MD simulations (Å)

Residues and Atoms	MD Distance to Zn ²⁺ (Å)		
	AfAglB-L-WT	AfAglB-L-ΔP1	AfAglB-L-ΔISP1
Asp47 Oδ1	2.2±0.2	2.2±0.2	2.3±0.4
Asp47 Oδ2	2.2±0.2	2.2±0.2	2.3±0.4
Asp161 Oδ1	2.2±0.2	2.3±0.3	2.4±0.5
Asp161 Oδ2	2.3±0.3	2.2±0.3	2.4±0.5
His163 Nε2	4.4±0.4	4.5±0.4	4.6±0.4
Glu360 Oε1	11.2±1.9	11.3±2.5	8.8±2.4
Glu360 Oε2	11.2±1.9	11.4±2.6	8.8±2.5

Supplementary data:

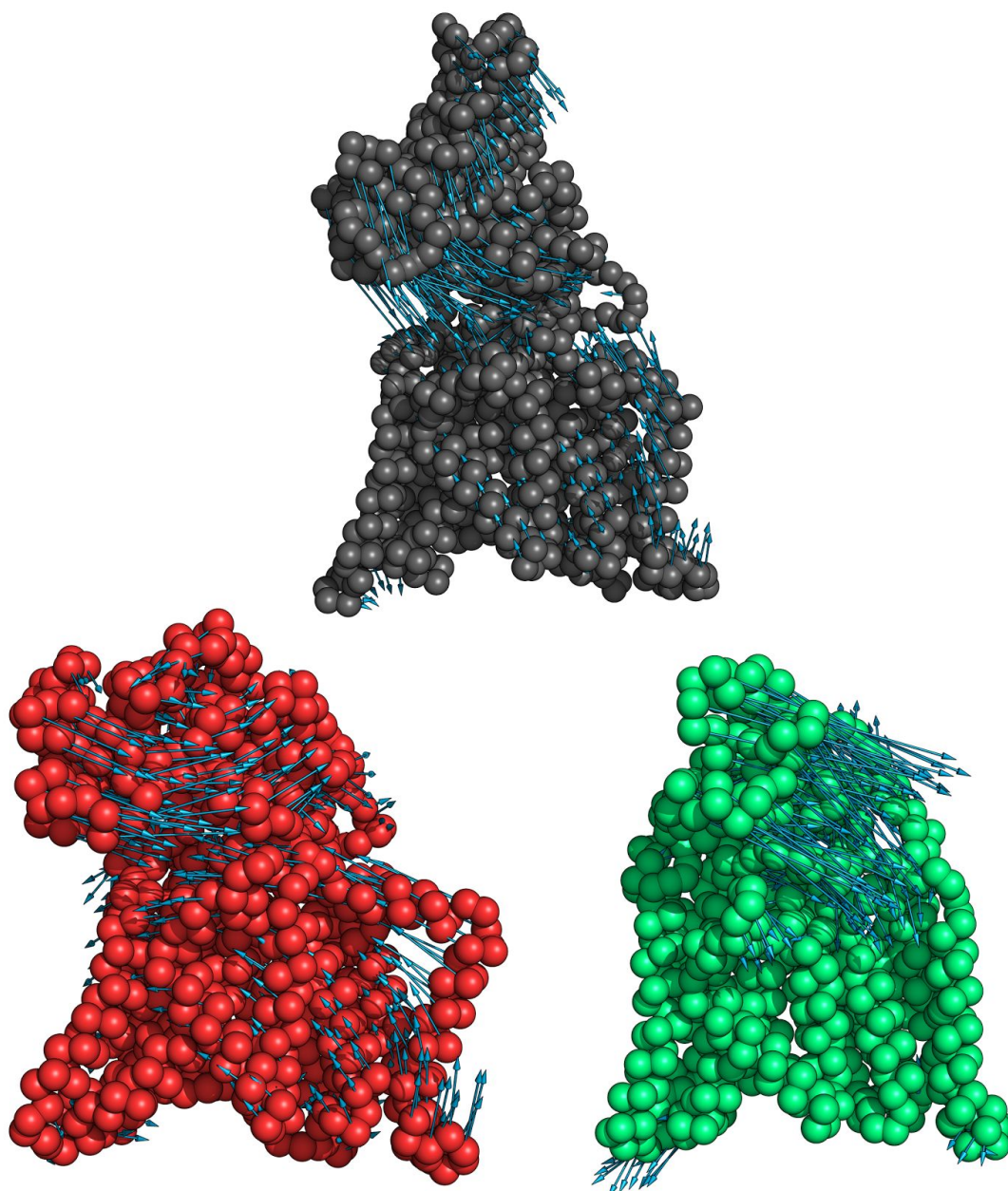


Figure S1. Porcupine plots depicting the movement described by the extreme projections of principal component 1 (PC1) of each group of replicates for the studied systems.

Highlights:

- Deletion of structural units causes instability and loss of secondary structure content.
- Conformational behavior of mutated systems deviate from the native structure.
- Coordination of the metal ion from the binding site is not disrupted.
- Hotspots of high fluctuation could be protected by engineering of chimeric OSTs.

5.3 Capítulo III

5.3.1 Preâmbulo

Esse capítulo é uma continuação dos estudos iniciados no capítulo I da tese, relativos à PglB. Aqui, o objetivo é a obtenção de uma estrutura do complexo completo da PglB, incluindo tanto o substrato proteico acceptor como o glicolípido doador. Em seguida, esse complexo foi avaliado por DM nos seus estados pré- e pós-catalíticos, gerando pontos de partida para futuros estudos que busquem a descrição do mecanismo catalítico envolvendo a ação dessa enzima por métodos QM/MM.

Durante a confecção dessa tese, simulações similares envolvendo esses estados foram publicadas [161] utilizando um campo de força (CHARMM) e uma membrana diferentes (POPC). Entretanto, o enfoque dado nessa publicação foi diverso, dando especial atenção à conformação adotada pelas hélices do domínio TM e da EL5. Na medida do possível, esses dados serão comparados àqueles já obtidos a fim de verificar possíveis convergências nas observações.

Até o presente momento, duas estruturas completas de OSTs são encontradas disponíveis em bancos de dados: a PglB de *C. lari* e a AglB de *A. fulgidus*. Os estados conformacionais cristalografados são: o cocrystalizado com um peptídeo acceptor (PglB) - cocrystalizado com um peptídeo acceptor covalentemente ligado na OST (AglB), o estado nativo (AglB) e o estado com um íon sulfato interagindo no sítio catalítico (AglB) - o qual, segundo os autores, mimetizaria um estado ligado ao glicolípido devido à presença de um fosfato em sua estrutura. Dessa forma, restam algumas lacunas em termos estruturais para a compreensão do ciclo de reação completo das OSTs. Nenhum estado contendo o substrato doador (LLO) foi obtido até o presente momento, de forma que não há indicação da conformação adotada pela enzima, nem dos respectivos substratos, neste estado.

5.3.2 Validação do modelo obtido para o LLO

Primeiramente, os estudos envolveram a obtenção de um modelo para o LLO. Em colaboração com o doutorando Pablo Ricardo Arantes do nosso grupo de pesquisas, realizamos a construção de uma molécula de LLO de *C. lari*, conforme descrito na seção de Procedimentos Metodológicos. Para a construção do modelo, um novo potencial diedral foi obtido (Figura 14), demonstrando concordância com o potencial obtido pela QM. Esse parâmetro foi validado utilizando um fragmento similar à unidade repetitiva da cauda hidrofóbica do LLO (geraniol) e comparando os valores encontrados na DM para duas propriedades, a densidade e a entalpia de vaporização. O valor para densidade

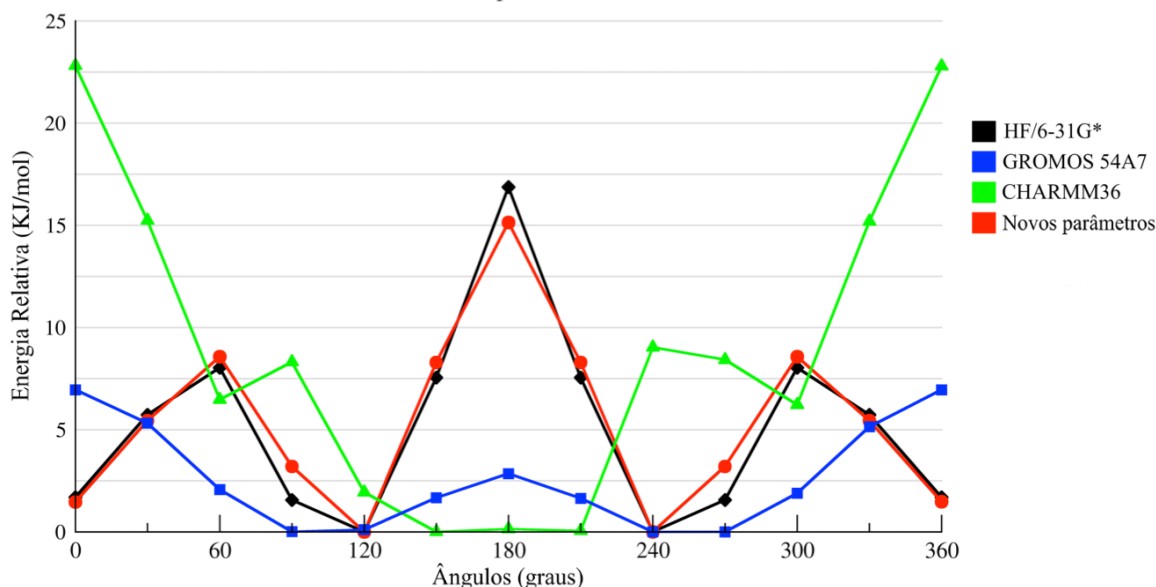


Figura 14 – Novo potencial ajustado (vermelho) ao dado quântico (preto). Os parâmetros antigos (azul e verde) não descreviam de forma adequada os mínimos de energia para essa torção.

correspondeu ao valor experimental ($0,89 \text{ g/cm}^3$), enquanto o valor da entalpia ficou próximo ($64,24 \pm 0,08 \text{ kJ/mol}$), com um erro menor que 10% em relação ao dado experimental ($58,83 \text{ kJ/mol}$).

A simulação do LLO inserido em membrana foi realizada por 1 microssegundo e indicou uma interação constante da glicana com a membrana (Figura 15), similar ao que foi verificado no trabalho de Kern e colaboradores [158]. Dessa forma, a glicana permaneceu em paralelo à membrana durante praticamente toda a simulação. Pelo uso de métodos de *clustering*, algumas conformações foram extraídas da DM para formação do complexo com a CIPglB. Realizando o atracamento molecular foi possível obter um modelo da OST bacteriana com todos os substratos (Figura 16) que respeitasse algumas distâncias pré-estabelecidas na literatura [158].

5.3.3 Dinâmica global dos domínios e unidades estruturais

Inicialmente, verificamos a variação estrutural ocorrida durante a simulação em relação à estrutura inicial de entrada (Figura 17). Foi possível verificar que a quantidade de variação estrutural geral da enzima foi comparável às simulações prévias observadas no Capítulo I (Figura 2 do artigo). Entretanto, analisando separadamente os domínios da proteína, percebe-se que o domínio TM apresentou um leve aumento quando comparado a esses dados previamente apresentados, possivelmente induzido pela presença do LLO. Não há grande discrepância na contribuição de cada domínio para a variação total da

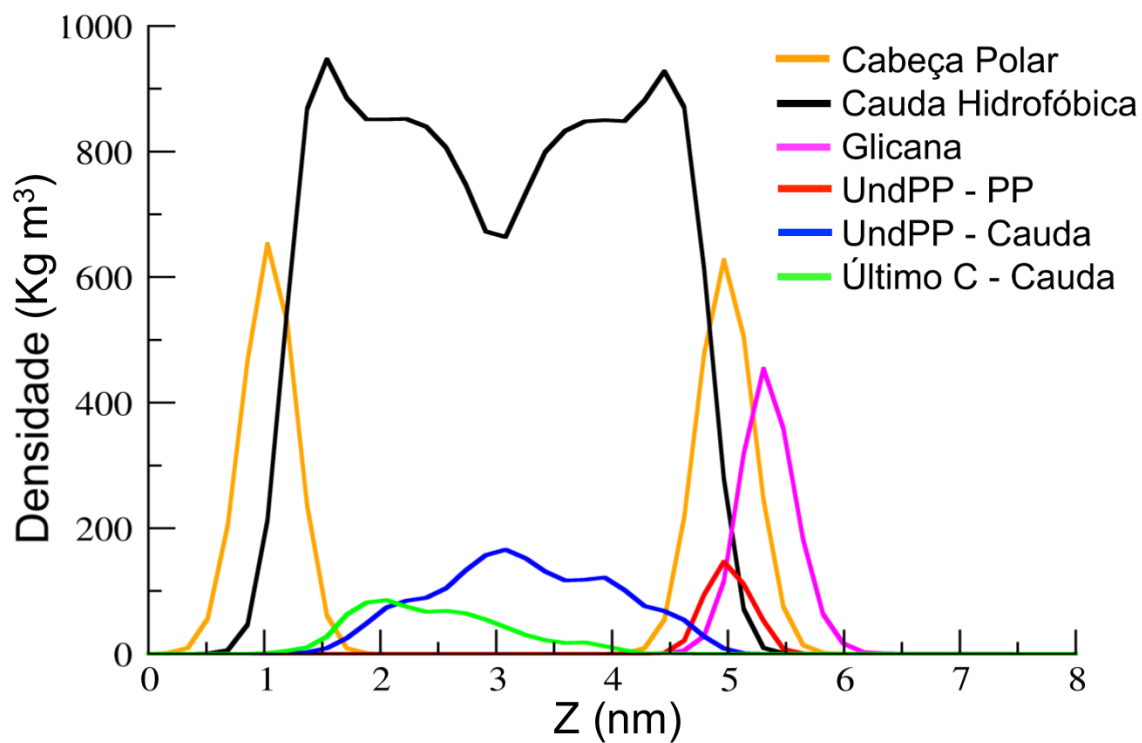


Figura 15 – Distribuição de densidade dos componentes do sistema simulado do LLO ao longo do eixo Z, demonstrando a posição em que se encontram esses elementos.

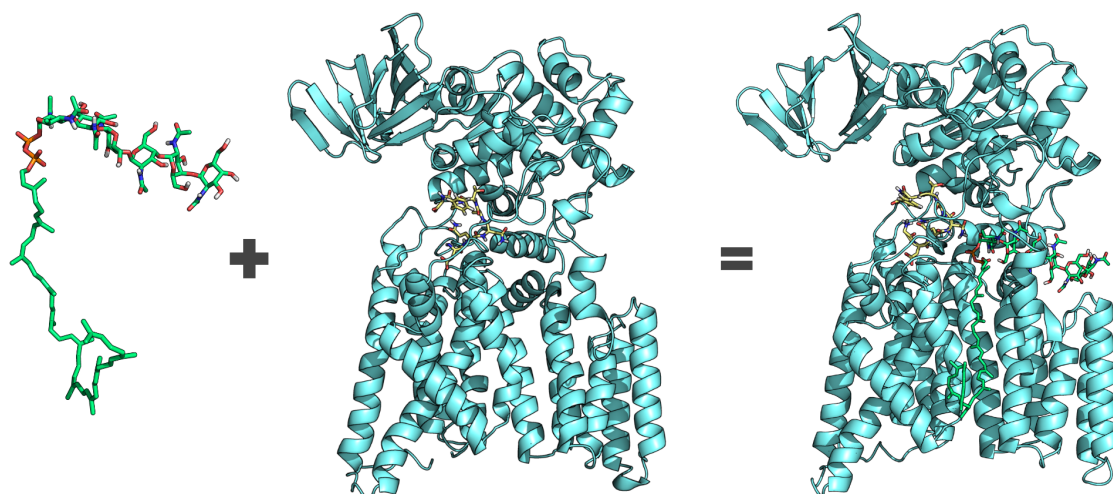


Figura 16 – Complexo ternário formado pela PglB e seus substratos, o peptídeo acceptor cocristalizado e uma molécula de LLO, modelada pelo nosso grupo.

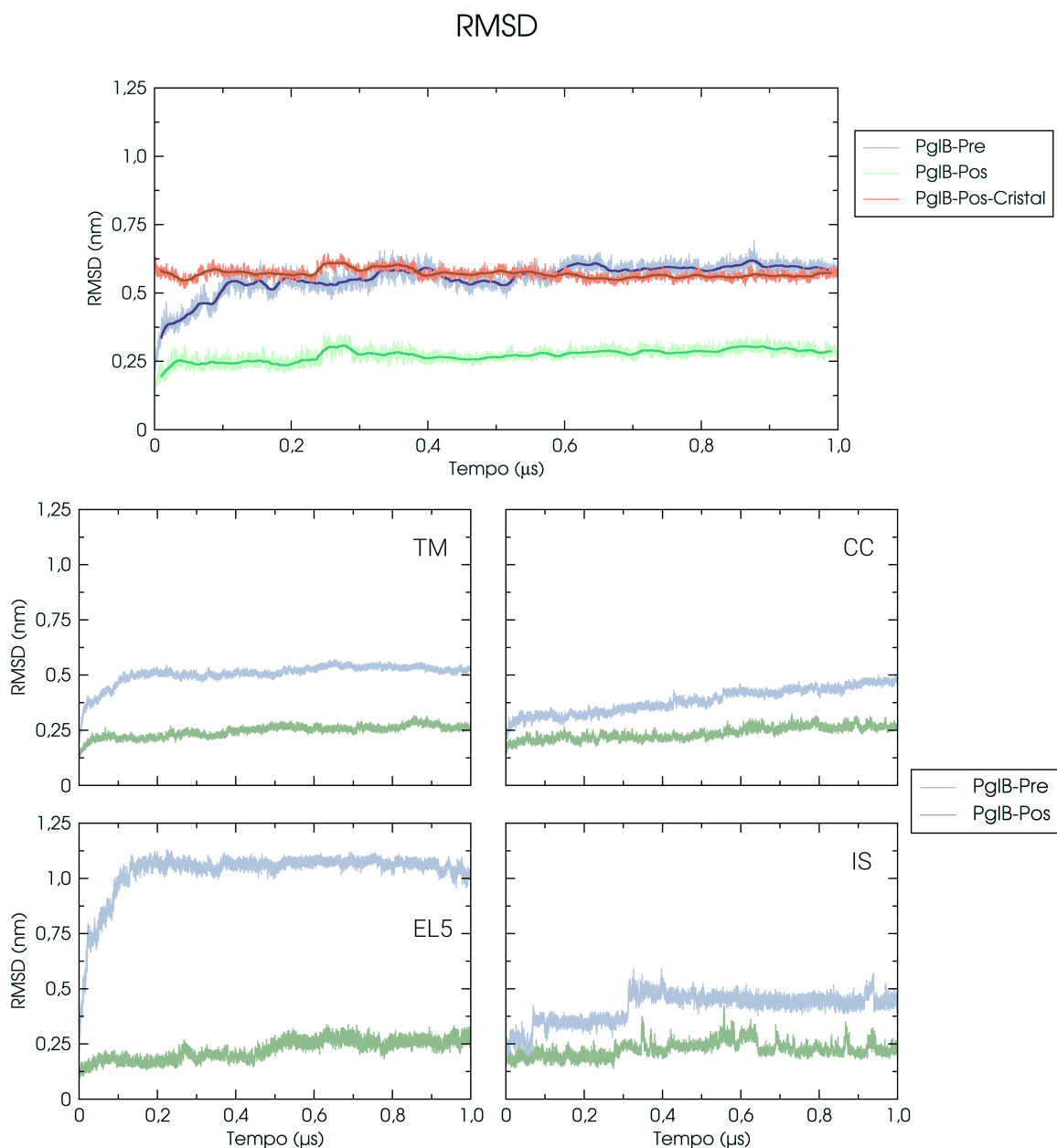


Figura 17 – RMSD total e separado por domínios da PglB. PglB-Pré = PglB no estado pré-catálise; PglB-Pós = PglB no estado pós-catálise; PglB-Pós-Cristal = PglB no estado pós-catálise utilizando como referência a estrutura inicial do pré-catálise

estrutura. Quanto às unidades estruturais CC e IS, que fazem parte do domínio PP, as duas variaram na mesma proporção (Figura 17) entre si. Entretanto, a unidade IS parece realizar rearranjos mais bruscos na sua estrutura em determinados momentos (Figura 17), reforçando a flexibilidade que é encontrada naturalmente nessa região para diferentes OSTs, como visto nos Capítulos I e II.

Um fator que parece contribuir para o domínio TM ser aquele que demonstra a

maior variação durante a simulação é a presença de suas alças externas, especialmente EL5 (Figura 17). O valor de RMSD para a EL5 chega a atingir cerca de 1 nm, o que possivelmente influencia nessa variação observada. Diferentemente dos casos estudados no Capítulo I, a EL5 mostrou-se ainda mais flexível e reorganizou-se mais intensamente na presença do LLO. O movimento observado na nossa simulação modificou consideravelmente a posição do motivo TYR-*plug* (Figura 18), especialmente do resíduo essencial Tyr293, porém de forma distinta do observado no trabalho recém publicado de Lee & Im [161], já que esses não observaram as mesmas interações feitas por esse resíduo. Ao analisar separadamente a região N-terminal e C-terminal dessa estrutura, a contribuição da primeira para a variação total foi o dobro da segunda (± 0.8 nm vs. ± 0.4 nm), concordando com o visto por Lee & Im. Essas informações reforçam a característica de alta flexibilidade da EL5, especialmente na porção N-terminal de sua estrutura, ao passo que auxilia a explicar a dificuldade em se obter dados tridimensionais cristalográficos para essa região. Ainda, é postulado na literatura [81] que essa porção N-terminal da EL5 (resíduos 283 à 307) assume uma estrutura secundária bem definida no processo de catálise. A proposta sugerida vem da estrutura cristalográfica da AglB de *A. fulgidus*, que possui uma pequena α -hélice. Durante a simulação, é possível verificar a formação de duas estruturas distintas: uma hélice- π que permanece formada por toda a simulação e carrega boa parte do TYR-*plug* e algumas estruturas β (folha, volta e dobramento) instáveis, que se formam e se desfazem diversas vezes.

A mutação da Tyr293 por um resíduo de alanina causa a redução da atividade catalítica da enzima em 7000 vezes [81], todavia sem afetar a ligação com o substrato peptídico. Já a troca desse resíduo por outro resíduo aromático, a fenilalanina, causa a diminuição de apenas 44 vezes na atividade. Esses dados indicam que não só a presença de um resíduo aromático e hidrofóbico é fundamental, mas também que o grupo hidroxila é relevante na função da Tyr293, possivelmente desempenhando um papel na catálise. Além disso, nesse mesmo trabalho prévio, foi verificado que esse resíduo não interage com os monossacarídeos da região distal da glicana, mas sim com a extremidade redutora. Na simulação incluída nessa tese foi identificado que o resíduo Tyr293 realizou dois tipos de interações com a N-acetilação em C4 do resíduo da bacilosamina (Figura 19): i) ligações de hidrogênio da hidroxila de sua cadeia lateral com o grupo cetona; ii) interações hidrofóbicas do anel aromático com o grupo metila. Dessa maneira, essa interação pode restringir a movimentação do monossacarídeo, permitindo uma otimização e manutenção de uma posição potencialmente adequada para o ataque nucleofílico da amida em C1.

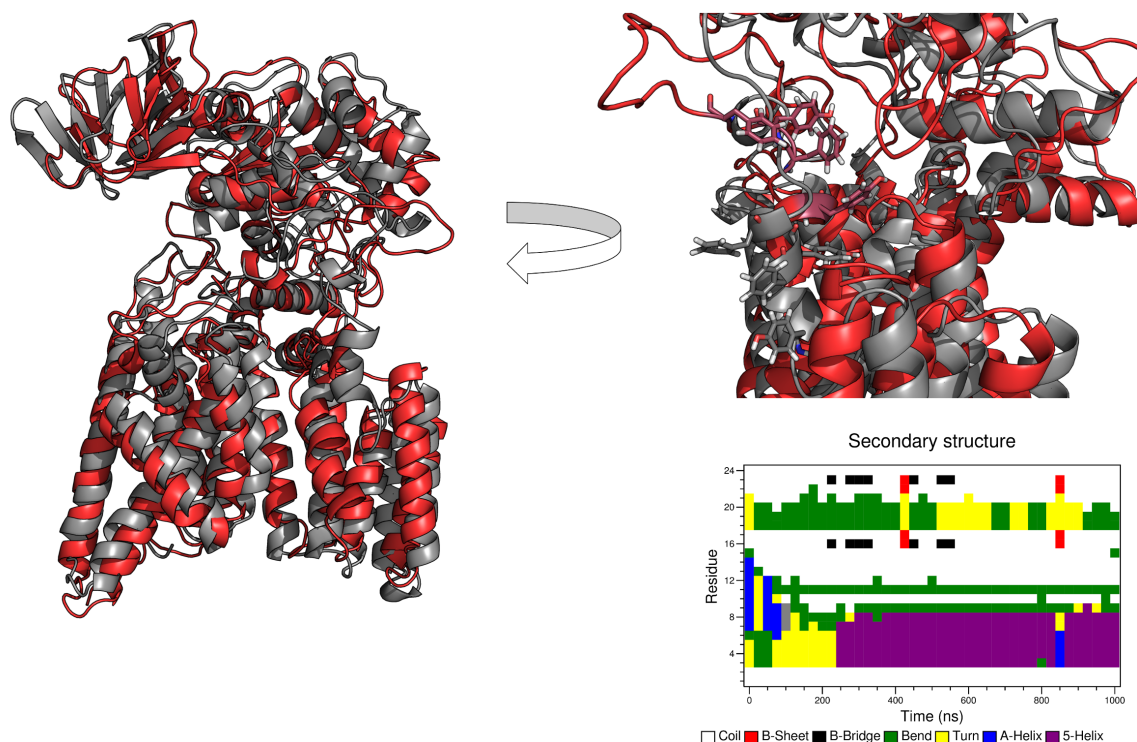
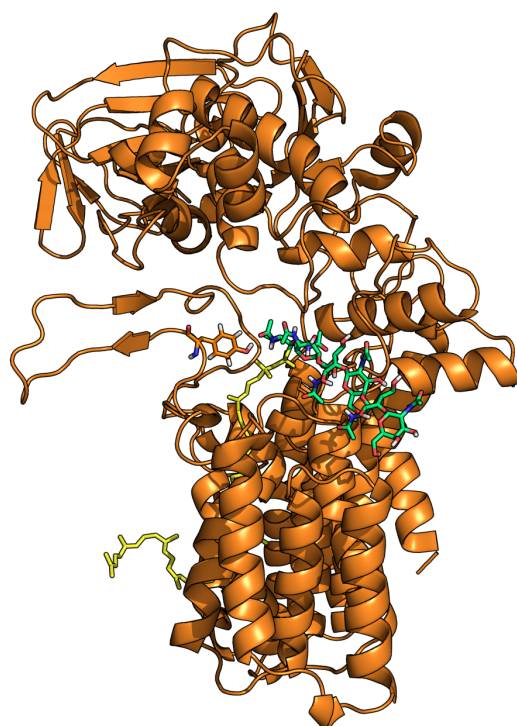


Figura 18 – Estrutura adotada pela EL5 durante a simulação: estado inicial (cinza); estado final (vermelho). *Acima, à direita.* Aproximação na cavidade da PglB, destacando o rearranjo do motivo Tyr-plug. *Abaixo, à direita.* Gráfico de tipo de estrutura secundária adotada pela região N-terminal da EL5 (resíduos 283-307) em função do tempo.

5.3.4 Coordenação da rede de interações no estado pré-catalítico

De modo semelhante ao visto no Capítulo I dos Resultados, faz-se relevante uma análise das distâncias envolvendo os resíduos do sítio catalítico, sendo eles o Asp56, o Asp154, o Asp156 e o Glu319, e o cátion divalente Mg^{2+} . Esses componentes são preditos como participantes de uma rede de interações que permitirá a transformação da amida em um bom nucleófilo a partir da alteração da configuração do átomo de nitrogênio. Portanto, as distâncias foram calculadas (Figura 20) buscando tanto avaliar o padrão da coordenação do metal como a presença de distâncias compatíveis com a existência de ligações de hidrogênio, fundamentais para a hipótese da amida torcida [61, 64] e para a reação catalítica (Figura 21). Os valores calculados vão ao encontro dos previamente descritos no artigo do Capítulo I, e posteriormente também verificados por Lee & Im, mostrando uma grande estabilidade na coordenação do metal pelos resíduos Asp56, Asp154 e Glu319. Mais uma vez, não foram vistas distâncias compatíveis com coordenação do metal para o resíduo Asp156, fato que discorda dos dados encontrados por Lee & Im, onde esse resíduo coordena o metal em diversas ocasiões. O papel do



Tyr293-OH-Bac-O9

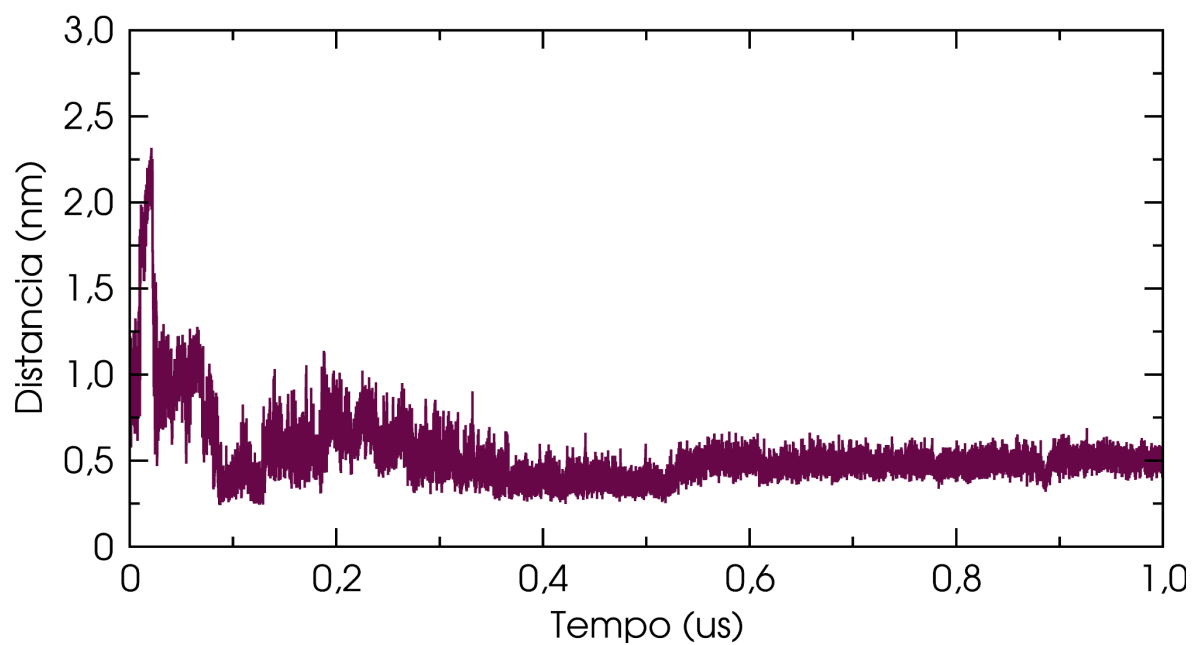


Figura 19 – *Acima*. Conformação de menor energia representativa de um estado onde a interação da Tyr293 de EL5 em relação à N-acetilação em C4 da Bac está bem caracterizada. *Abaixo*. Distância entre o grupo -OH da Tyr293 e o O9 da Bac em função do tempo.

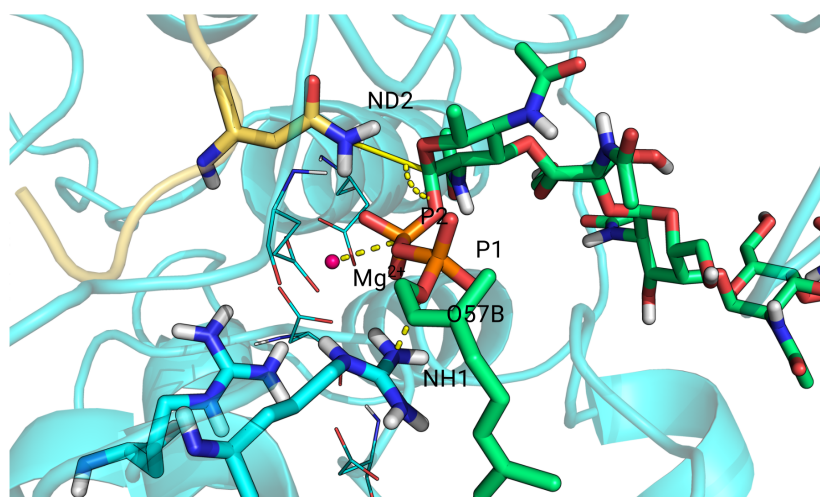
Átomos	Distância para Mg ²⁺	Distância para Asn-Nδ2
Asp56-Oδ1	2,1±0,01 Å	3,5±0,2 Å
Asp56-Oδ2	2,1±0,01 Å	5,2±0,4 Å
Asp154-Oδ1	2,2±0,01 Å	-
Asp154-Oδ2	2,1±0,01 Å	-
Glu319-Oε1	2,1±0,01 Å	5,1±0,3 Å
Glu319-Oε2	2,1±0,01 Å	3,6±0,3 Å

Figura 20 – Distâncias dos resíduos catalíticos para o metal divalente e para o nitrogênio (ND2) da Asn.

resíduo Asp156, e, por consequência, do resíduo na terceira posição do motivo DXD ainda permanece sem uma explicação clara, apesar de estudos mostrarem ser essencial para a atividade das GTases [80]. As distâncias iniciais, que guiaram o atracamento molecular, foram mantidas, indicando a estabilidade dos substratos no estado pré-catálise e a existência de condições apropriadas para a reação de transferência da glicana.

5.3.5 Caracterização do comportamento estrutural e conformacional da PglB

Para melhor compreender como a dinâmica da enzima é afetada por sua ação, antes e após a transferência da cadeia de oligossacarídeos, empregamos a análise de componentes principais (ACP). Buscando obter estados conformacionais relevantes para o mecanismo catalítico, foram analisados os dois primeiros componentes principais para os sistemas pré e pós-catalítico, tanto separadamente (Figuras 22 e 23), como considerados dentro de um mesmo subespaço (Figura 24). Analisando o comportamento separadamente para a simulação do estado pré-catalítico, o comportamento global da proteína aponta inicialmente para uma flutuação por mínimos locais, até que na metade da simulação ela encontra uma região de menor energia, ali permanecendo até o final. Entretanto, se a análise for realizada separadamente para os diferentes elementos estruturais, vemos padrões específicos para cada região da enzima: o domínio



Átomos	Distância na DM ^a	Ângulo na DM
Mg ²⁺ - P2-LLO	2.86 Å ± 0.01 Å (< 6 Å)	-
Asn-ND2 - C1-Bac	3.53 Å ± 0.2 Å (< 4 Å)	-
Arg375-NH1 - P1-LLO	4.0 Å ± 0.3 Å (< 6 Å)	-
Asn-ND2 - C1-Bac - O57B-LLO	-	109.1° ± 8.6° (< 90°)

^aValores em parênteses representam os esperados para uma reação catalítica SN2

Figura 21 – *Acima*. Representação esquemática das interações essenciais para a ocorrência da reação catalítica da PglB. *Abaixo*. Distâncias e ângulo observados na DM para essas interações. Entre parênteses, os valores esperados experimentalmente para esse processo.

TM (excluindo EL5) move-se gradualmente durante a simulação alcançando um mínimo mais estável apenas no fim da simulação; a EL5 comporta-se de forma semelhante, porém atinge um mínimo antes, por volta do mesmo período que a proteína como um todo se estabiliza (metade da simulação); a unidade CC move-se intensamente e apesar de encontrar uma região de menor energia na porção final da simulação, continua explorando o espaço conformacional; por fim, a unidade IS, assim como visto no gráfico de RMSD, faz mudanças de estados conformacionais bruscas, porém encontra um mínimo após o primeiro terço da simulação e ali permanece até o fim, sendo a estrutura que mais rapidamente estabilizou-se no cálculo realizado.

A simulação do estado pós-catalítico funcionou quase como uma continuação da simulação do estado pré-catalítico, porém começando de uma conformação extraída no tempo 618 ns. Foram caracterizadas algumas variações dentro do perfil conformacional

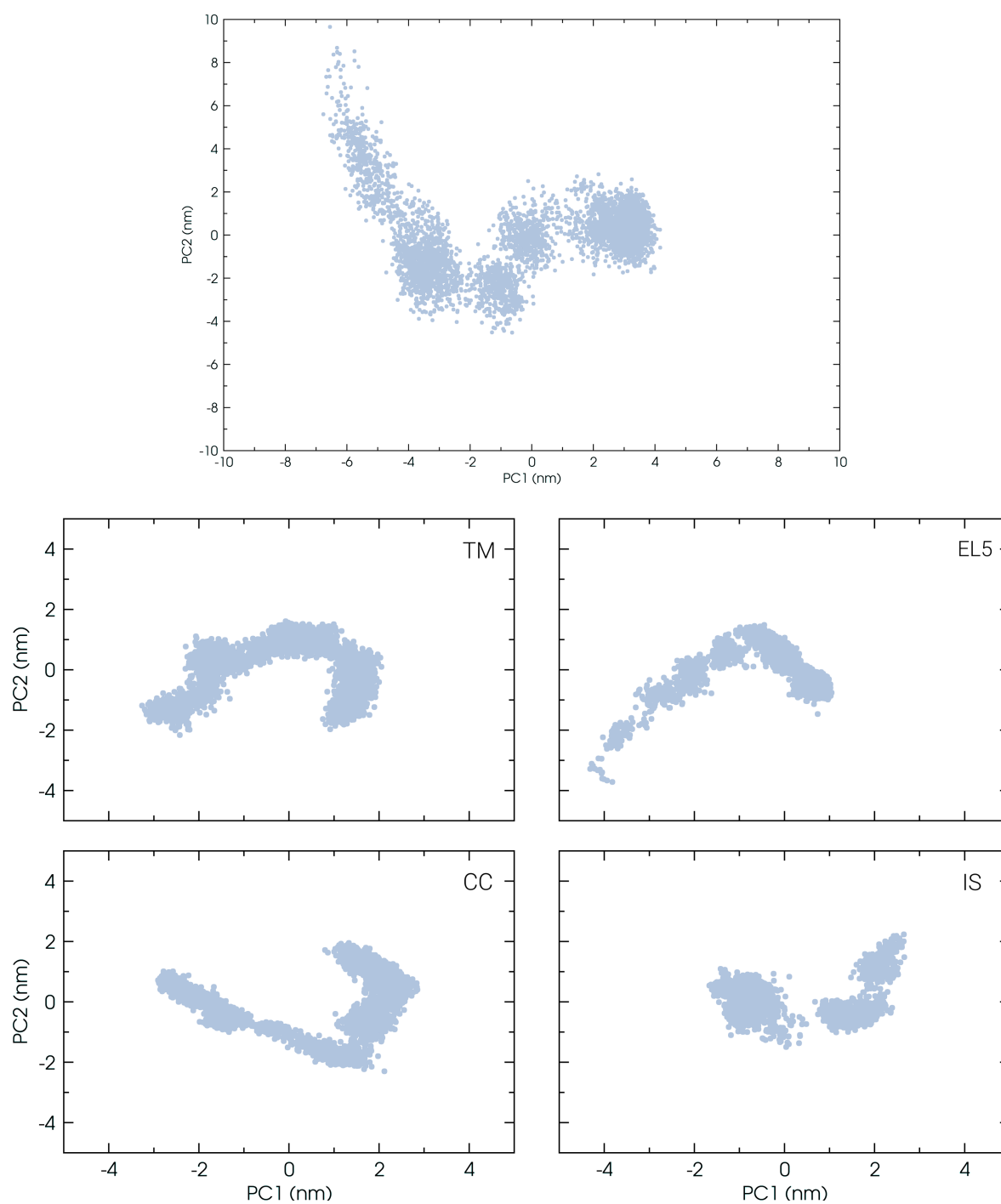


Figura 22 – Projeções dos dois primeiros componentes principais da PglB em seu estado pré-catalítico inteira e separada por seus domínios e unidades estruturais: TM (sem EL5); EL5; CC; IS.

da PglB descrito pelos CPs do estado pós-catalítico, possivelmente influenciados pela mudança entre os substratos: o domínio TM continuou a exploração conformacional, sendo a única estrutura a não se estabilizar nessa simulação; a EL5 sofre uma mudança para um mínimo de energia próximo ao seu estado inicial; a unidade CC encontra uma conformação e chega a interconverter para o seu estado inicial; a unidade IS manteve-se no mínimo inicial e explora as regiões próximas, sem alterar seu perfil. Esse dado indica que após a transição do estado pré-catalítico para o pós, a região com maior mobilidade é o domínio TM, que inclusive permanece explorando o subespaço formado pelos dois primeiros CPs sem se estabilizar, possivelmente devido ao novo substrato ali presente (UndPP sem a glicana).

Buscando uma comparação direta entre os perfis conformacionais dos dois estados, as duas trajetórias das simulações foram concatenadas e a APC foi feita em cima dessa nova trajetória, garantindo assim que os dados gerados pela análise sejam comparados em um mesmo subespaço. No domínio TM (novamente sem EL5), percebe-se que há um período com sobreposição inicialmente, todavia a conformação das hélices transmembranas é modificada pela troca de estados e substratos. Na EL5 há uma semelhança bem maior, com um pequeno desvio conformacional mais para o fim da simulação do estado pós, indicando pequenas alterações nos movimentos dessa subestrutura. O caso da unidade CC é mais curioso: a conformação extraída da simulação do pré-catalítico, apesar de ser a de menor energia globalmente, não pertence ao mesmo mínimo da porção final da simulação do estado pré-catalítico. Ao ser simulada no estado pós-catalítico, ela não segue o mesmo caminho da simulação do pré-catalítico, selecionando essa região de mínimo de forma mais estável e ampliando a amostragem desse fosso, como visto na interconversão de estados para a APC no parágrafo anterior. Assim, a conformação dessa região parece ter sido estabilizada nesse ponto pela troca da glicana, anteriormente ligada à UndPP, para o peptídeo. Por fim, o estado IS parece ser o único totalmente convergido entre as duas simulações, pois toda a área amostrada no tempo de simulação do pós-catalítico é sobreposta à conformação final do estado pré-catalítico, indicando a identificação de um mínimo global entre os dois cálculos. Essa informação faz sentido, já que o ponto de troca entre substratos está relativamente distante dessa unidade e a mudança não parece ter efeitos drásticos na estrutura da enzima.

5.3.6 Análise de estados conformacionais da PglB

Com a aplicação da APC, estados conformacionais de menor energia e, portanto, possivelmente relevantes dentro da dinâmica da proteína estudada, puderam ser

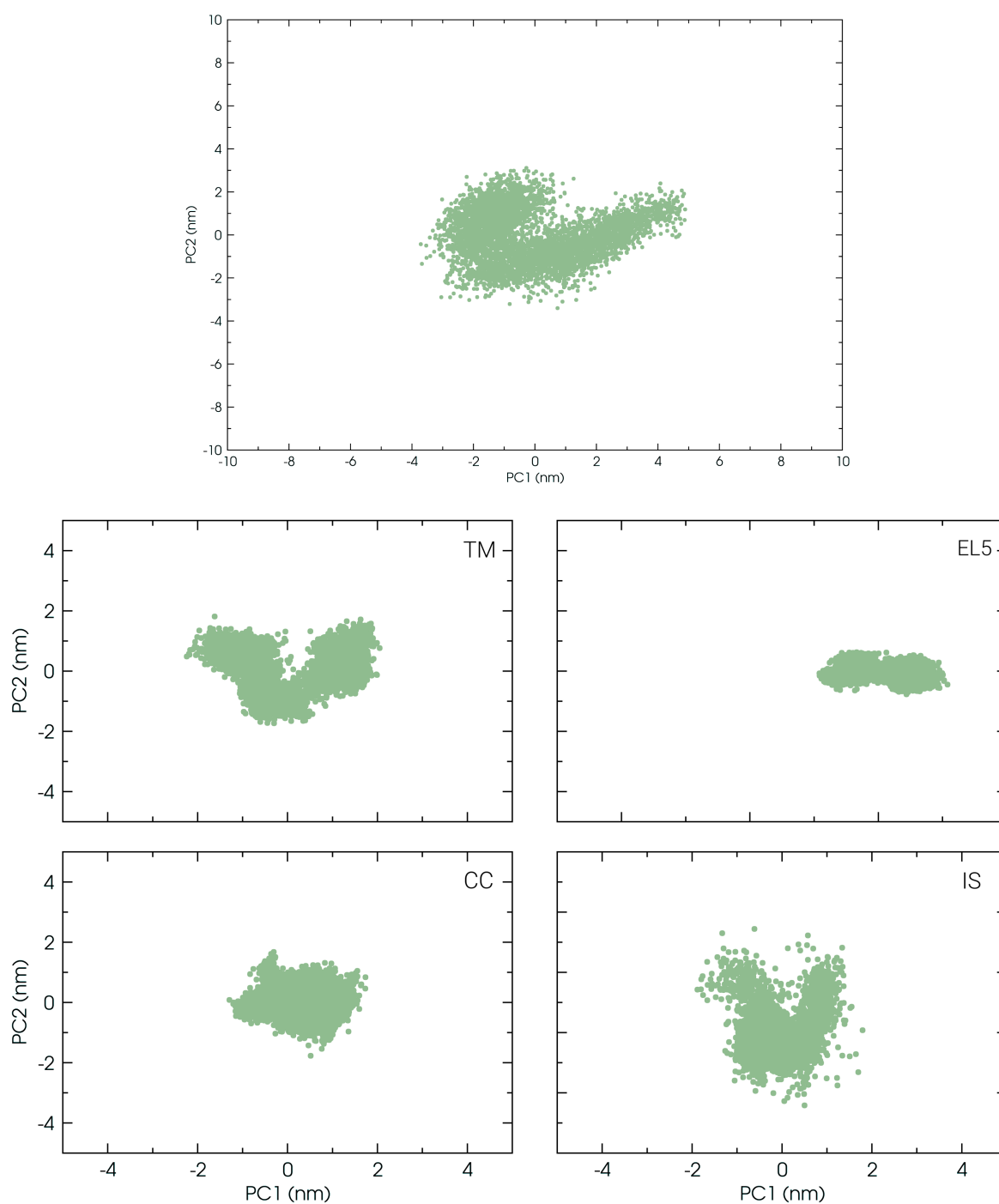


Figura 23 – Projeções dos dois primeiros componentes principais da PglB no estado pós-catalítico inteira e separada por seus domínios e unidades estruturais: TM (sem EL5); EL5; CC; IS.

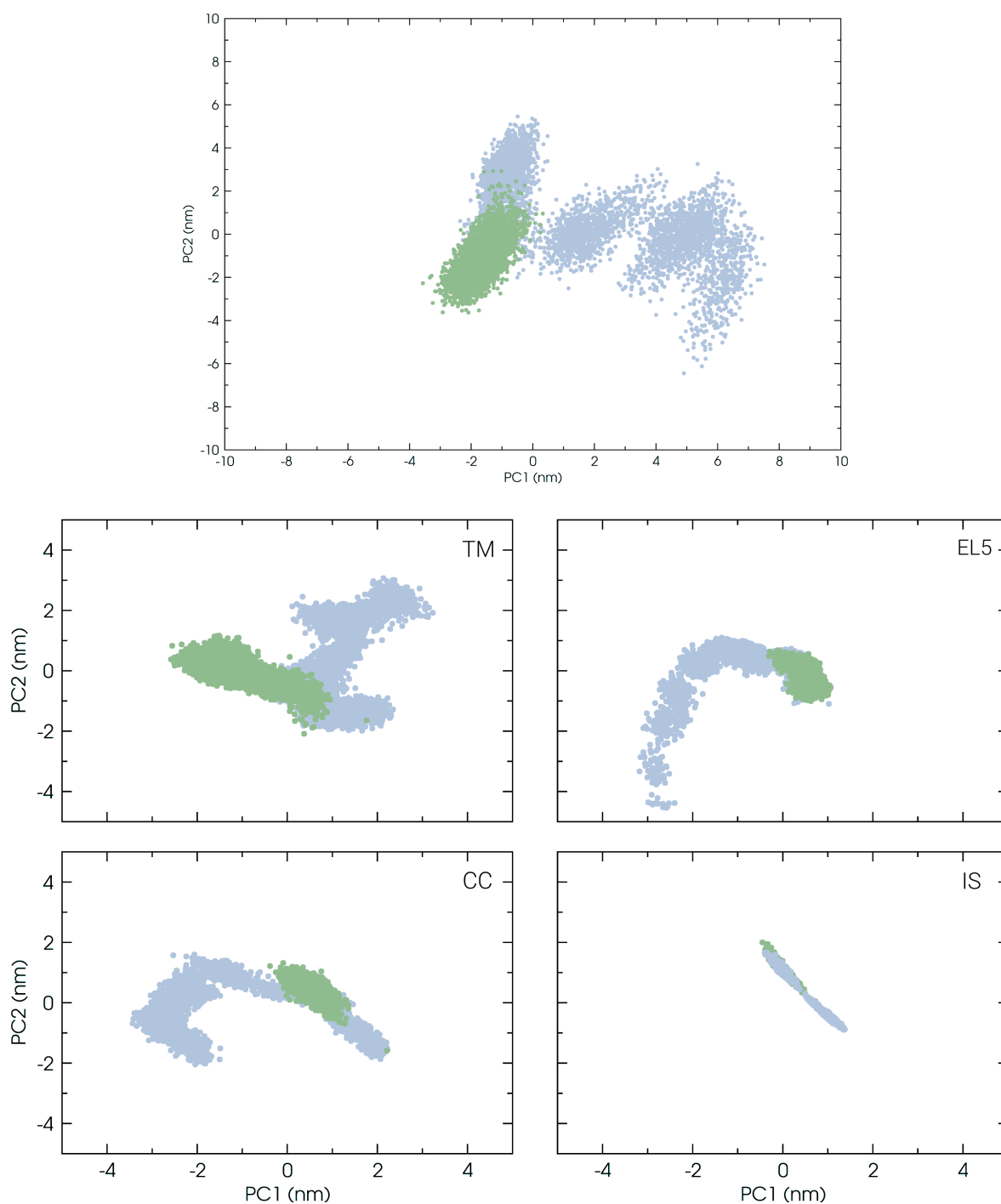


Figura 24 – Projeções dos dois primeiros componentes principais da PglB em seus dois estados para um mesmo subespaço, tanto inteiras como separadas por seus domínios e por suas unidades estruturais: TM (sem EL5); EL5; CC; IS.

selecionados. Foi realizada, então, uma comparação com o estado inicial da PglB que foi empregado na simulação do estado pré-catalítico. Para essa comparação, a análise de *porcupine plot* ou gráfico porco-espinho foi utilizada. Nessa análise, dois estados diferentes são comparados e flechas são geradas conforme o tamanho da variação entre as diversas regiões da molécula de interesse. A partir desse gráfico, caracterizou-se tridimensionalmente a variação encontrada em decorrência da introdução do LLO na estrutura da enzima, destacando-se um movimento pronunciado nas hélices transmembrana localizadas na região à direita do agrupamento de hélices, próximo ao N-terminal da EL5 (Figura 25). Esse movimento intenso na presença do substrato doador também foi visto no trabalho de Lee & Im, onde duas conformações, aberta e fechada, foram propostas para essa região. No trabalho dessa tese movimentos intensos das hélices TM5-9 empurram a porção N-terminal da EL5, causando o rearranjo dessa área e explicando a grande flexibilidade verificada durante a simulação. O próprio motivo Tyr-*plug* é empurrado para cima, sendo aproximado por um tempo em direção à bacilosamina. Calculando as distâncias identificadas no trabalho de Lee & Im (Figura 25), entre os carbonos γ dos resíduos Leu202 e Leu365 e entre os carbonos α dos resíduos Tyr293 e Met368, foi identificado que os valores são semelhantes, o que indica que a EL5 está na sua conformação aberta, reforçando esse dado prévio.

Para o estado pós-catalítico, as mesmas distâncias foram calculadas (Figura 26) mostrando que a enzima permanece com a EL5 na conformação aberta, devido à interação com a undecaprenila pirofosfatada que permanece interagindo no sulco hidrofóbico após a glicosilação do peptídeo. Utilizando a mesma análise de *porcupine plot*, entretanto agora com os estados pré-catalítico e pós-catalítico de menor energia, percebem-se diferenças muito sutis: um movimento para fora da membrana da hélice TM9, que liga na porção N-terminal da EL5, bem como um movimento discreto dessa mesma região da EL5 em direção ao corpo da proteína. Ademais, alguns pequenos movimentos são vistos nas unidades CC e IS, que não parecem influenciar fortemente na direção de um novo estado conformacional da PglB, endossando o que foi observado nos gráficos dos CPs.

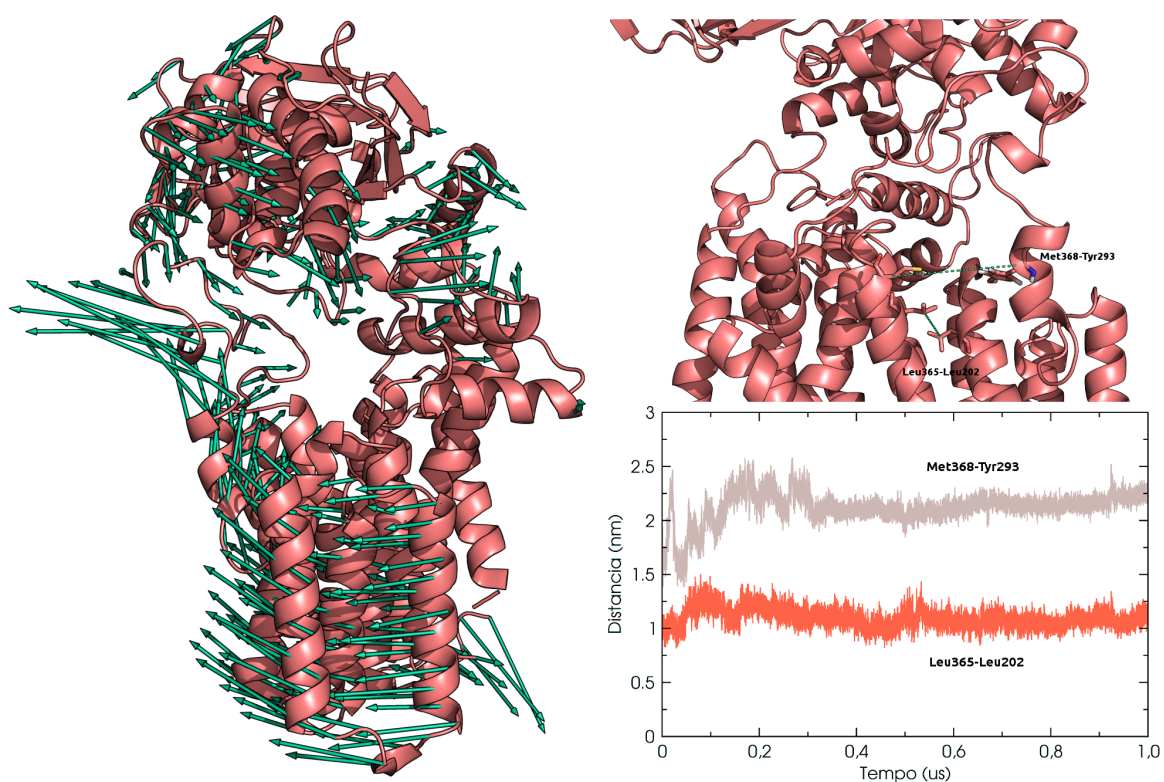


Figura 25 – *Porcupine plot* diferenciando o estado inicial e a conformação de menor energia identificada por APC na simulação do estado pré-catálise. *Direita*. Distâncias entre os resíduos identificados para a conformação aberta de EL5 e do sulco hidrofóbico onde o LLO interage.

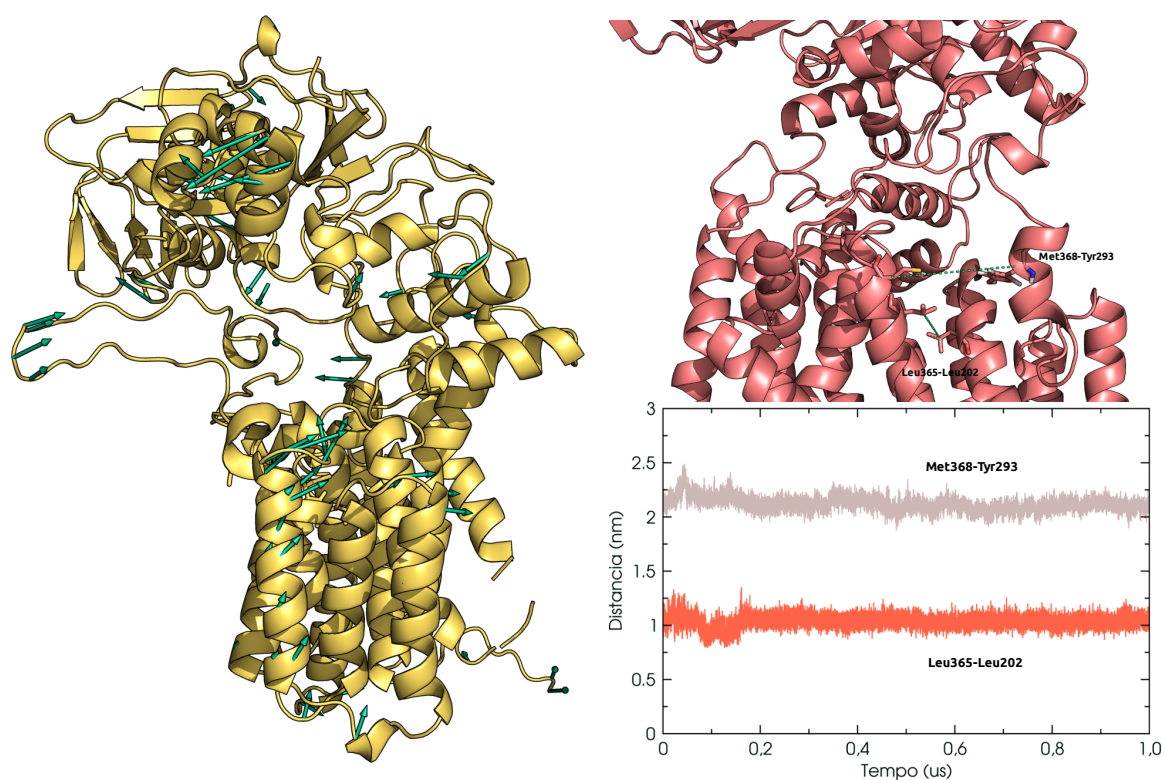


Figura 26 – *Porcupine plot* diferenciando o estado inicial do sistema pós-catálise e a conformação de menor energia identificada por APC para essa mesma simulação. *Direita*. Distâncias entre os resíduos identificados para a conformação aberta de EL5 e do sulco hidrofóbico onde o LLO interage.

6 Discussão geral

*“Nothing in Glycobiology makes sense,
except in the light of evolution”*

Ajit Varki

O Grupo de Bioinformática Estrutural tem se envolvido desde seu início com a área de glicobiologia. O foco inicial do grupo envolveu pequenas estruturas sacarídicas, como di- e oligossacarídeos, buscando acessar a conformação dessas biomoléculas [204–206]. Com o avanço do poder computacional, problemas mais complexos, como glicoproteínas inteiras, começaram a ser abordados [129, 207, 208]. Desse movimento em direção a problemas maiores e de magnitude mais ampla, surge a necessidade de buscar um melhor entendimento das bases moleculares relativas à construção dessas estruturas glicoproteicas. Em 2011, com a obtenção da primeira estrutura cristalográfica completa de uma OST, foi possível iniciar os primeiros estudos de caracterização da dinâmica dessa enzima. Durante o meu Mestrado, iniciado em 2012, começamos a analisar essa família de enzimas empregando DM para avaliação da conformação e dinâmica da PglB, bem como ferramentas de filogenia para tentar compreender a sua evolução molecular [159].

Nosso estudo envolvendo a caracterização da dinâmica estrutural da PglB foi o primeiro estudo com resolução atômica desse tipo [160]. Partindo da estrutura cristalográfica, propomos um primeiro modelo para a região N-terminal da EL5 já que, devido à grande flexibilidade dessa região, ela não pôde ser obtida no estudo cristalográfico original. Por intermédio de simulações por DM, verificamos que a EL5 assume conformações distintas durante a simulação, confirmando a flexibilidade intrínseca postulada para essa estrutura. Buscamos avaliar também a PglB em estados não descritos no experimento cristalográfico, como a PglB não complexada e na presença do glicopeptídeo produto da reação. Contudo, o principal enfoque desse trabalho foi estudar a coordenação do metal divalente do sítio ativo, bem como o comportamento das interações de resíduos conservados e relevantes para o processo da catálise. O mecanismo de ação proposto para a PglB foi inicialmente postulado juntamente a sua estrutura cristalográfico [64]. Posteriormente, foi refinado por um segundo trabalho [61], onde a hipótese da carboxamida torcida ganhou maior suporte. No nosso trabalho caracterizamos a coordenação do íon metálico por MM e QM, extraindo janelas da simulação onde a Asn aceptora possuía diferentes conformações, bem como os resíduos

que coordenam o íon catalítico. Dessa forma, verificamos que em uma das conformações do sítio ocorreu uma mudança na configuração do nitrogênio da amida, a partir da influência causada pelas ligações de hidrogênio dos resíduos catalíticos, reforçando a proposta mecanística da literatura. Ainda, havia sido sugerido que 3 ou 4 moléculas de água participariam da coordenação metálica [64]. Com os nossos cálculos postulamos que, na ausência da undecaprenila pirofosfatada, apenas uma molécula de água assume essa função, enquanto os outros resíduos (Asp56, Asp154 e Glu319) variam entre coordenações mono- e bidentadas.

Esse trabalho sofreu com alguns obstáculos. A PglB é uma molécula consideravelmente grande para estudos computacionais (711 AA) sendo, ainda, uma proteína transmembrana. Esses fatores levaram à construção de um sistema possuindo uma quantidade próxima aos 200.000 átomos, exigindo um alto custo computacional. Por ser um sistema complexo e de grande magnitude, idealmente seria necessário realizar simulações mais longas ou um maior número de replicatas para cada sistema. Por consequência, apesar de ser um trabalho iniciado no Mestrado, o tema das OSTs demonstrou a exigência de uma maior dedicação, sendo assim continuado no Doutorado, paralelamente aos estudos da outra OST, a AglB, que teve sua estrutura cristalografada descrita em meados de 2013, pouco antes do início do meu Doutorado. Soma-se a isso o fato de que durante o desenvolvimento inicial desse trabalho diversos questionamentos surgiram, tais como: Qual o papel das unidades estruturais das diferentes OSTs? Qual o comportamento da PglB quando na presença de todos seus substratos? Podemos extrapolar esses dados para outras OSTs e empregar em estudos relacionados à gli-coengenharia? Delinearam-se assim as diversas vertentes que foram analisadas nesse trabalho.

As unidades estruturais das OSTs constituem uma questão particular. Essas enzimas, apesar de serem amplamente distribuídas nos domínios Eukarya e Archaea, além de terem uma considerável presença no domínio Bacteria, são bastante diferentes quando se fala em termos de estrutura primária, ou seja, sequência de aminoácidos. Apesar disso, analisando as duas estruturas 3D atualmente disponíveis, percebe-se claramente uma semelhança topológica nos seus respectivos enovelamentos globais, apesar de possuírem variações locais nas estruturas [32], talvez um indicativo de que ocorreu convergência evolutiva. O mesmo pode ser observado para as estruturas parciais obtidas até o momento [62, 63, 65, 71, 209]. A estrutura considerada mais distante evolutivamente e diferente em termos estruturais é a única estrutura parcial do domínio Eukarya, descrevendo o enovelamento o domínio C-terminal da subunidade STT3 de *S. cerevisiae* como um agrupamento de α -hélices (α -helix bundle), enquanto nas estruturas procarióticas caracterizadas esse enovelamento é do misto α/β (*mixed- α/β*). Como

comentado anteriormente na Introdução, enquanto já se identificou diferentes unidades estruturais (IS, P1, P2) em organismos procarióticos, em eucariotos apenas foi vista a unidade CC. Dois fatores devem ser levados em consideração: a ausência de mais estruturas eucarióticas, dificultando a noção de distribuição de unidades nesse domínio, e o fato de que a subunidade STT3 faz parte de um complexo OT nesse domínio, que em geral possui 5-8 subunidades. Uma hipótese da literatura [65] é de que algumas dessas subunidades poderiam desempenhar o papel das unidades que foram encontradas nos domínios procarióticos. Entretanto, a baixa identidade de sequência é um contratempo que torna difícil a identificação de similaridade entre as proteínas desses domínios. Um trabalho [210] com análise evolutiva da via de N-glicosilação (que será mais discutido a seguir) tentou encontrar ortólogos nos diferentes domínios, usando tanto identidade de sequência como busca por padrões (HMM), porém também não obteve sucesso.

Uma abordagem tentada, entre o meu Mestrado e o meu Doutorado, foi uma análise de aspectos filogenéticos das OSTs, empregando o método de Máxima Verossimilhança [211] e procurando entender como ocorreu a evolução molecular dessa enzima, assim como verificando a distribuição das unidades estruturais pelos domínios (Figura 27). A baixa identidade e as grandes diferenças nos tamanhos das enzimas (como exemplo, a Af-AglB-S1 e a S2 são as menores OSTs já identificadas e essas possuem cerca de 550 resíduos, enquanto a OST de *Thermoplasma acidophilum* possui cerca de 1500 resíduos) foram condições que prejudicaram inicialmente a obtenção de resultados confiáveis. Posteriormente, em 2016, um artigo [210] com análises filogenômicas foi publicado, porém para várias GTases importantes da via de N-glicosilação, entre elas as OSTs (Figura 28). Para tal fim, o autor empregou o método de Inferência Bayesiana [212], conhecido por aplicar a probabilidade posterior como parâmetro de confiança. Os dados obtidos por Lombard e colaboradores têm similaridades importantes com o que verificamos. Um exemplo claro disso é que, em seu trabalho, ele observou que é bastante provável que o domínio Bacteria tenha recebido a OST por transferência horizontal de genes (THG) do domínio Archaea, em específico vindo do filo *Euryarchaeota*, algo que também se verifica na árvore proposta por nós. O ramo ligado aos eucariotos estaria próximo ao superfilo putativo TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota), chamado de Proteoarchaeota ou Proteoarchaea. Esse último dado é, inclusive, reforçado por dados recentes que mostram que a N-glicana do organismo *Pyrobaculum calidifontis* é similar à do domínio Eukarya. Portanto, há dados que apontam para uma divisão mais clara de como ocorreu a distribuição de OSTs pelos domínios da vida. Devido à grande sobreposição de dados entre os trabalhos, decidimos descontinuar a nossa investigação.

Considerando as informações disponíveis, partimos para uma tentativa de realizar

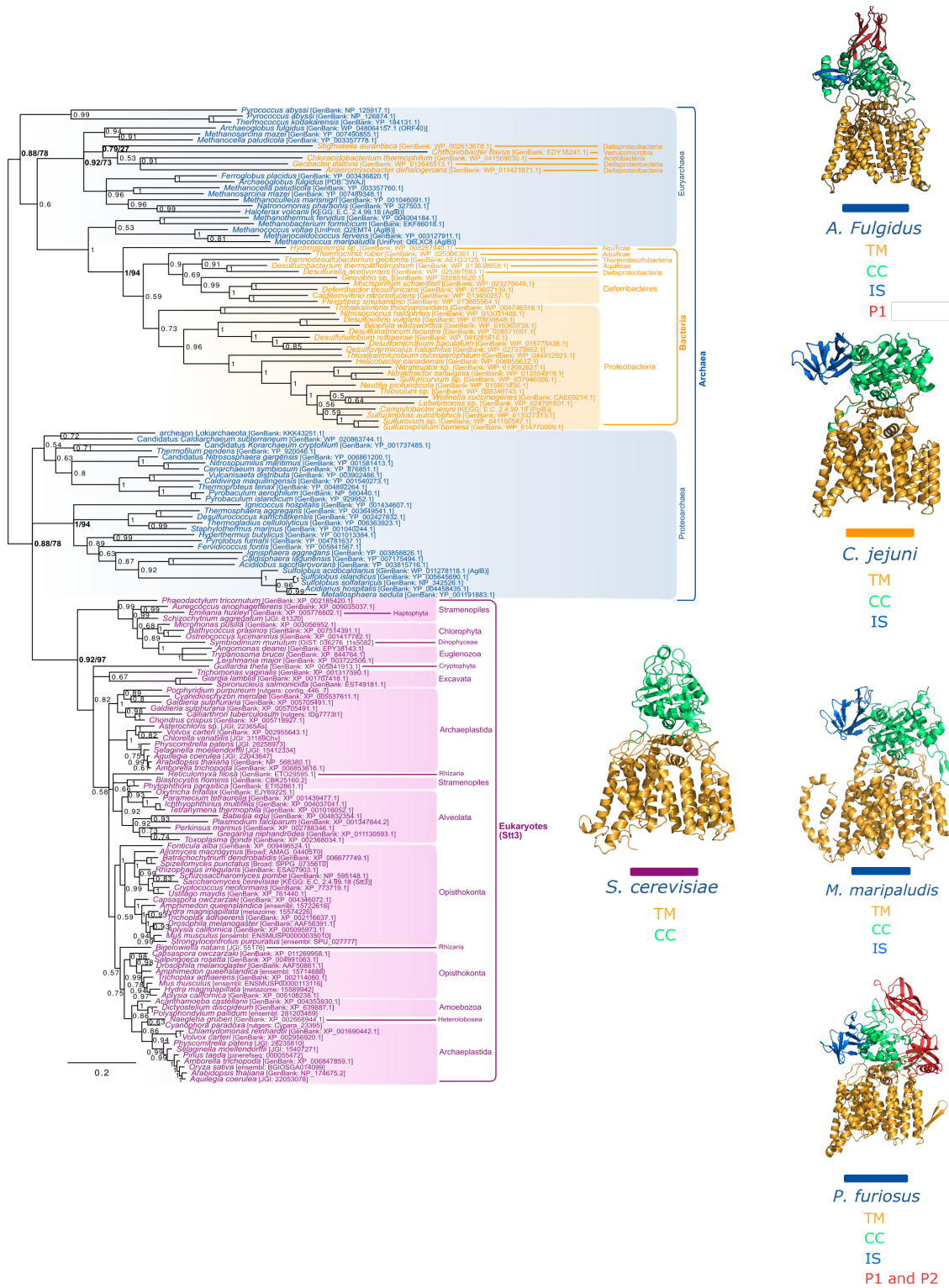


Figura 28 – Árvore filogenética gerada pelo método de Inferência Bayesiana para 163 sequências de aminoácidos de OSTs [210]. Os domínios seguem o código de cores, enquanto em uma escala de 0 a 1 são mostrados os valores de probabilidade posterior que indicam a confiança do dado (onde 1 é o mais alto). As estruturas modeladas por nós demonstram as respectivas composições de unidades estruturais para espécies representativas.

STT3 completa desse organismo é possível, pois as topologias transmembrana das OSTs são provavelmente similares para todos domínios da vida [215]. Como esperado, dado o tamanho da sequência e os dados prévios obtidos para a estrutura parcial, apenas uma unidade estrutural foi encontrada, a CC. Na tentativa de modelar a enzima de organismos eucarióticos que estejam mais próximos do domínio Archaea, a OST de *Trichomonas vaginalis* foi a obtida com maior confiança (96% dos resíduos modelados com confiança >90%). Mesmo assim, sua estrutura não é muito informativa e talvez não seja confiável: parece possuir uma pequena organização estrutural posicionada na região da unidade IS com algumas pequenas β -fitas e com algumas voltas que não interagem com o restante do corpo da enzima. Esse caso ajuda a ilustrar as dificuldades na elucidação estrutural de OSTs, onde mesmo na presença de estruturas com bons indicadores para modelagem comparativa (100% de cobertura da sequência, identidade acima de 30%), não é possível obter para uma descrição acurada dos casos encontrados nessa classe de enzimas.

Em Archaea, a modelagem torna-se ainda mais complicada: a despeito de possuir algumas estruturas parciais descrevendo o domínio C-terminal e 3 estruturas descrevendo enzimas completas (*A. fulgidus*), não é possível obter modelos para a maior parte das espécies. Quando tratamos dos filos dentro do superfilo Proteoarchaeota, a situação é ainda mais complicada: os domínios C-terminais tem estruturas com confiança baixíssima, geralmente resultando em modelos fisicamente irrelevantes. Nesse domínio da vida, as estruturas de OSTs são quase caso-específicas, tamanha a diferença entre as enzimas das espécies. Uma hipótese seria de que a pressão evolutiva é mais alta nesses organismos, devido às condições extremas, o que implicaria em alterações mais frequentes e perda de similaridade nas espécies atuais. Na atual conjuntura, somente obtendo estruturas cristalográficas completas correspondendo a cada filo poderíamos avaliar a diversidade de unidades estruturais e suas atribuições, sejam elas estruturais ou para a função das OSTs.

Apesar da diversidade de estruturas encontradas nessas enzimas, muito pouco ainda se sabe sobre a importância funcional das suas unidades estruturais. Nesse sentido, os resultados apresentados no Capítulo II dessa tese são os primeiros dados voltados para a elucidação da dinâmicas dessas subestruturas e como elas afetam as OSTs. Com isso, foi verificado que, para a AgIB-L de *A. fulgidus*, as unidades P1 e IS são, no mínimo, importantes para a integridade estrutural da OST. Os resultados apontam para a relevância dessas estruturas no processo de manutenção da estrutura em altas temperaturas, já que observamos que a estrutura nativa se desorganiza, assim como a interface de interação com os substratos é alterada e ocorre perda de estrutura secundária em alguns casos, indicando um possível início de desnaturação. Quanto

ao sítio catalítico, esse manteve-se com a coordenação estável em todos os sistemas mutados, assim como visto para as simulações dos sistemas nativos. Isso de certa forma não foi uma surpresa já que, como mostrado no artigo do Capítulo II, um estudo prévio demonstrava a total funcionalidade da enzima após a deleção da unidade IS em *P. furiosus*, sugerindo assim que a coordenação do íon não era necessariamente afetada *in vitro* [65]. Entretanto, na AglB-L de *P. furiosus* há a presença de duas unidades periféricas P1 e P2 que envolvem toda a unidade CC, com P1 ainda fazendo contatos com a unidade IS, proporcionando uma estabilidade bem maior em regiões da OST que demonstraram intensa flutuação tanto nesse artigo como no trabalho do Capítulo I. Essas duas unidades periféricas, especificamente, parecem ser mais ligadas à proteção estrutural contra fatores externos, como as altas temperaturas e as altas concentrações salinas, já que são ricas em β -fitas e com menor mobilidade. Ainda, até o momento, só foram encontradas em organismos do domínio Archaea, que comumente habitam ambientes inóspitos sob as condições mais variadas.

O conhecimento de que algumas dessas unidades parecem possuir uma certa independência estrutural e que algumas evidências apontavam para a hipótese da termoestabilidade levou-nos a fazer algumas tentativas envolvendo engenharia de proteínas com a construção de uma PglB quimérica, visando a atingir a estabilidade em altas temperaturas para a enzima bacteriana. Realizamos duas simulações dessas quimeras na temperatura ótima de *A. fulgidus*, cada uma com um método diferente de inserção: a primeira empregando o software Modeller [216] diretamente e selecionando o modelo com melhor porcentagem de resíduos em regiões favoráveis no gráfico de Ramachandram [217]; a segunda realizando uma substituição molecular, alinhando as estruturas da AfAglB-L com a CIPglB e removendo a estrutura de arquêia. Contudo, em 50 ns de simulação nenhum dos sistemas manteve-se estável o suficiente, obtendo-se resultados levemente melhores no segundo caso, com a estrutura da unidade P1 na PglB mantendo-se mais parecida com a original AglB e realizando mais interações com o resto do corpo da enzima (Figura 29). As regiões C-terminais das estruturas das PglBs quiméricas movimentaram-se em direção à membrana, modificando o comportamento das enzimas nativas. Por essas razões, concluímos que mais estudos ainda são necessários para o desenvolvimento de uma nova série de enzimas termoestáveis. Uma proposição que fazemos é a de realizar mutações na PglB para que a região de interface com a nova unidade transferida fique com um ambiente físico-químico mais semelhante ao da AglB (Figura 29). Futuramente, uma segunda geração de PglBs quiméricas será testada em condições extremas para buscarmos a formação de OSTs bacterianas termoestáveis.

O mecanismo completo das OSTs vem sendo estudado [61], em conjunto com o objetivo de obter-se estruturas dessas enzimas em todos os estados do ciclo de reação.

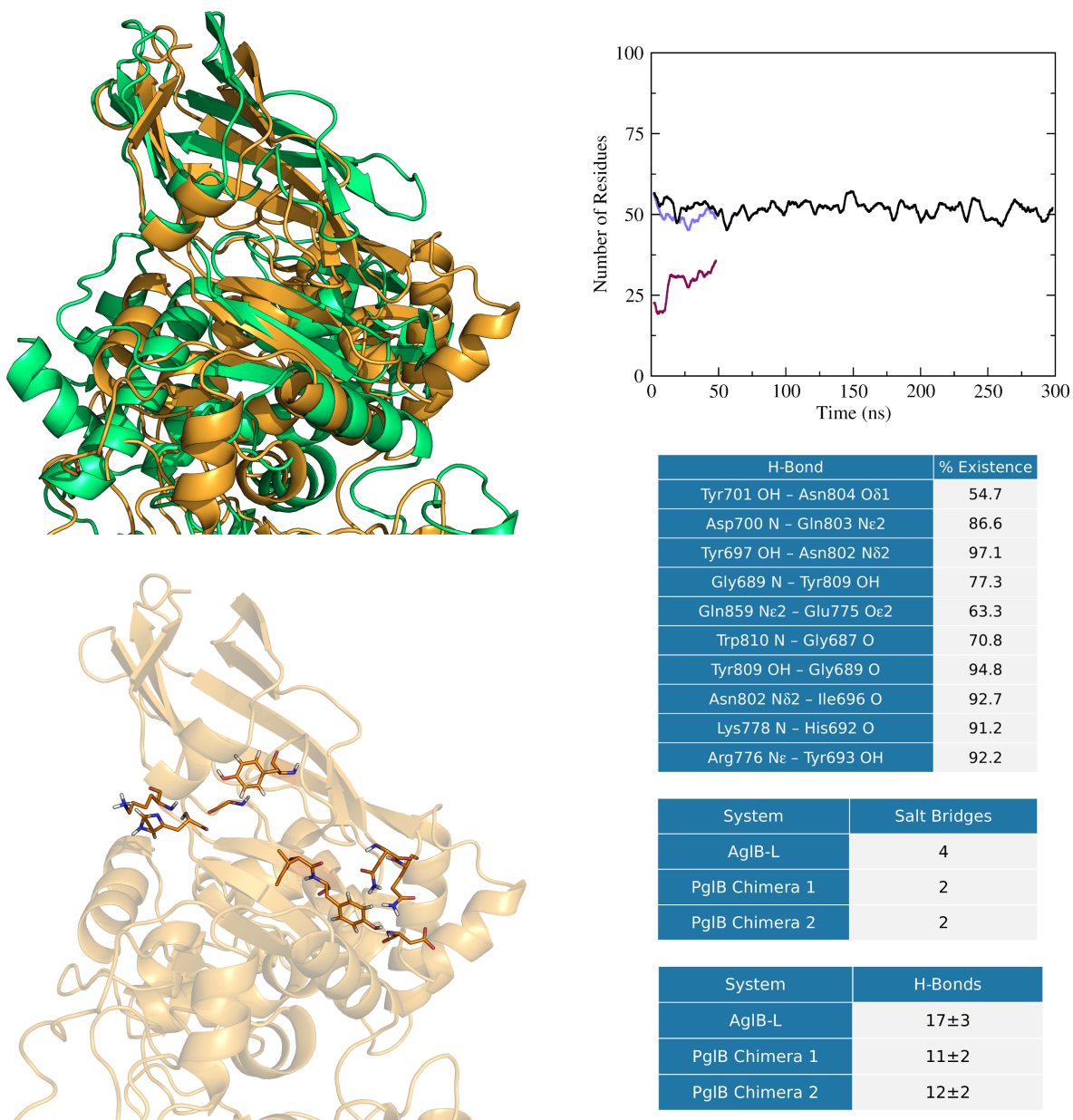


Figura 29 – Análise de estabilidade estrutural das PglBs quiméricas e interações realizadas durante as simulações de cada uma.

Atualmente, o ciclo de reação foi proposto por Matsumoto e colaboradores (Figura 30) [66] e envolve os seguintes passos: 1) o estado nativo ou apo é representado pela estrutura cristalográfica da AfAglB-L forma 2 (ID 3WAK), com a EL5 ordenada e o Glu360 não acoplado ao sítio catalítico; 2) o estado com o substrato proteico é representado pela CIPglB (ID 3RCE), com as interações da Asn e da Thr da assinatura bem estáveis no sítio e a formação da portinhola por onde ocorrerá a catálise; 3) estado do complexo ternário, sem estrutura obtida experimentalmente; 4) estado transiente após a reação de transferência com o reagente e o produto ainda dentro da enzima, também sem dados experimentais. Contudo, foi visto [70] que o motivo WWD é flexível o suficiente para que ocorra a liberação da Ser/Thr da assinatura e a saída do produto da reação; 5) o estado ligado apenas ao carreador lipídico, após a saída da glicoproteína é representado pela estrutura da AfAglB-L forma 1 (ID 3WAJ), com um íon sulfato representando o grupamento fosfato do carreador, com a EL5 completamente desordenada. O recém aceito trabalho de Lee & Im [161] propõe atualizações para os passos faltantes do ciclo: no passo 3, a forte ligação do LLO com a Arg375 induz movimentos nas hélices TM, especialmente na região que compreende as hélices TM5-9, o que aumenta ainda mais a flexibilidade da porção N-terminal de EL5, causando a conformação aberta e permitindo a glicosilação; no passo 4, a undecaprenila pirofosfatada mantém-se ligada à proteína, enquanto o N-terminal da EL5 permanece desordenada, até que a glicoproteína sai da enzima e a EL5 torna-se totalmente desordenada. Os resultados que incluímos nessa tese tem sobreposição com os dados de Lee & Im. Todavia, trazemos algumas observações novas para esse tema. Vimos algo similar em relação ao movimento das hélices do domínio TM, porém demonstramos que isso causa um rearranjo da EL5, distanciando-a da cavidade da enzima e permitindo a flexibilidade característica dessa região nesse estado. Vimos ainda que o Tyr-*plug* assume conformações distintas em nossas simulações, realizando inclusive interações entre Tyr293 e a N-acetilação na posição C4 da Bacilosamina, algo que não havia sido observado na literatura previamente e que, como sugerido por Lizak e colaboradores [81], pode auxiliar na restrição de um estado conformacional mais adequado à catálise. Esses dados ainda servirão como um ponto de partida, fornecendo conformações relevantes para estudos de QM buscando descrever o mecanismo catalítico da PglB.

Um ponto que permanece não esclarecido é a N-acetilação obrigatória na posição C2 do primeiro monossacarídeo em Bacteria e Eukarya. Após todos esses estudos, ainda não parece haver nenhuma explicação para esse requerimento. Dentro da cavidade da PglB, esse grupo acetoamida não apresenta nenhuma interação com resíduos próximos, descartando, inicialmente, algum papel específico na estabilização de uma conformação pré-catalítica. O LLO permaneceu dentro da cavidade como observado por Lee & Im e

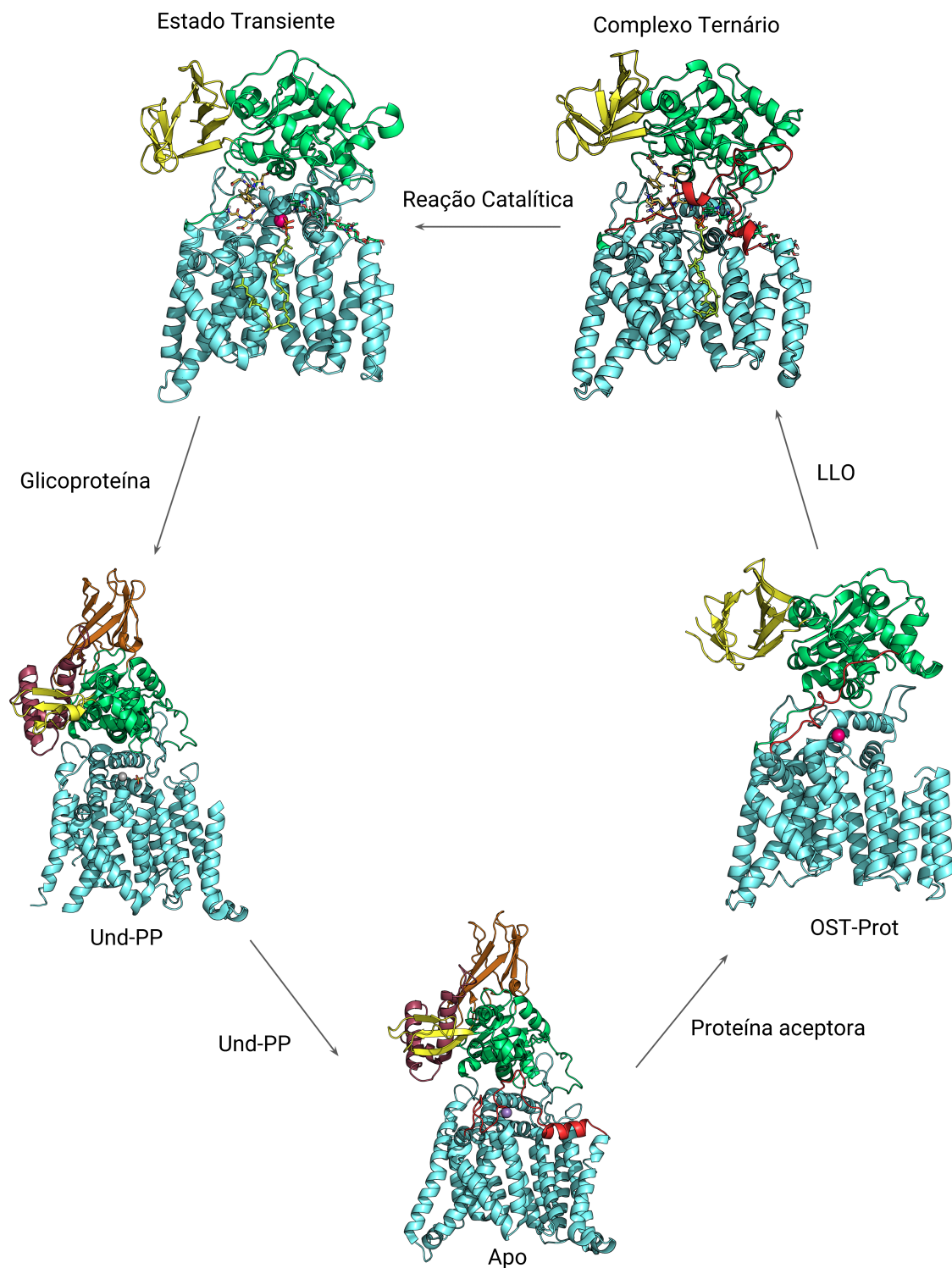


Figura 30 – Ciclo de reação das OSTs indicando cada estado e seus respectivos substratos e produtos, baseado no que foi proposto por Matsumoto e colaboradores [66]. Os estados Apo, OST-Prot e Und-PP foram obtidos por cristalografia, enquanto o Complexo Ternário e o Estado Transiente foram obtidos por DM.

também nos nossos dados; as interações tanto de um fosfato com o íon metálico, como do outro fosfato com a Arg375, restringem a movimentação do substrato doador. Assim, fazemos duas propostas nessa tese a respeito da importância desse grupo funcional:

- a presença desse grupo pode modificar a flexibilidade das ligações glicosídicas no LLO, gerando um estado conformacional que permite (na presença) ou não (na ausência da acetilação) a entrada na OST. No trabalho de caracterização dos LLOs em colaboração com o colega Pablo Arantes verificamos, por DM, que a presença de uma glicose no lugar da bacilosamina pode alterar os valores do ângulo ψ da ligação glicosídica com o carreador lipídico, o que poderia gerar um novo estado conformacional incompatível com o sítio da PglB. A análise de um LLO de Archaea, mais especificamente o de *A. fulgidus* ou de outro Euryarchaeota, poderá auxiliar na resposta a essa pergunta, pois os organismos desse filo não possuem a obrigatoriedade dessa N-acetilação (nessa espécie, particularmente, é um resíduo de glicose). Estas análises encontram-se em andamento;
- a ausência das interações realizadas pela N-acetilação nessa posição, sejam elas intra ou intermoleculares, altera o perfil conformacional populado por essa molécula e impossibilita a entrada do LLO na cavidade de ligação. Duas abordagens que podem auxiliar a resolução dessa questão futuramente seriam a metadinâmica ou o *pulling* seguido de *umbrella sampling* para forçarmos a entrada do substrato na cavidade e verificarmos os aspectos essenciais inerentes ao processo.

Não podemos descartar também a possibilidade dos nossos experimentos não terem descrito apropriadamente as interações no interior da enzima, impedindo de caracterizarmos corretamente a ligação com a OST.

Por fim, um aspecto que foi bastante abordado nesse trabalho foi a coordenação no sítio catalítico. Particularmente, o papel do resíduo na terceira posição do motivo DXD (DXH na AfAglB) foi discutido nos Capítulos II e III como não sendo muito claro, pois apenas na estrutura cristalina de AfAglB ele participava na coordenação. Realizamos uma simulação da PglB em seu estado apo por 1 microssegundo também, porém ela não foi abordada nessa tese. Apesar disso achamos relevante mencionar que, durante essa simulação, a Asp156 coordenou o metal de forma bidentada (2,05 Å e 2,12 Å de médias) juntamente com o Asp56, o Asp154 e o Glu319 nos últimos 200 ns da simulação. Isso reforça a hipótese de que esse resíduo pode realizar um papel de auxílio na manutenção do metal preso e estável no sítio catalítico. Para tanto, na ausência de elementos que facilitam a coordenação, como o peptídeo e o LLO com seu pirofosfato, ele participa na coordenação do íon divalente. Inclusive, o íon parece estar

sempre presente no sítio [61], sendo provável que desempenhe um papel estrutural e de organização de todo o entorno do sítio catalítico nessa enzima. Ressalte-se, novamente, que a MM não é a ferramenta mais adequada para descrever coordenação metálica, portanto devemos ter cuidado ao interpretar os dados advindos dessas simulações.

7 Conclusões

A presente tese tratou dos objetivos traçados e permitiu:

I. Descrição da dinâmica do sítio catalítico da PglB de *C. lari*:

- i. Descrever a coordenação catalítica no sítio ativo por dois métodos distintos (MM e QM), demonstrando como o comportamento dos resíduos pode influenciar no mecanismo;
- ii. Demonstrar que a presença dos substratos proteico e glicoproteico influenciam na movimentação das unidades estruturais e na dinâmica conformacional da EL5, estrutura essencial no processo catalítico.

II. Avaliação do papel funcional das unidades estruturais da AglB de *A. fulgidus*:

- i. Caracterizar a alteração estrutural causada pelas mutações, com perda considerável do perfil conformacional nativo da enzima, possivelmente induzindo à diminuição ou perda total da atividade;
- ii. Verificar se há manutenção dos requisitos básicos para catálise em todos sistemas;
- iii. Realizar uma primeira geração de OSTs quiméricas e identificar interações que poderão ser empregadas em uma segunda geração de enzimas com termoestabilidade aumentada.

III. Caracterização dos estados pré- e pós-catalíticos da PglB de *C. lari*:

- i. Construir um modelo validado para o LLO e identificar que a sua interação com a PglB afeta diretamente o comportamento de uma região do domínio TM e da EL5, particularmente no motivo Tyr-*plug*, gerando novas interações entre enzima e substrato que podem interferir na catálise;
- ii. Obter estados conformacionais para os sistemas pré- e pós-catálise, a partir de regiões de menor energia durante as simulações, incluindo estruturas com sobreposição conformacional para serem empregados em cálculos de QM envolvendo o mecanismo catalítico da PglB.

8 Perspectivas

O desenvolvimento dos trabalhos contidos nessa tese trouxe novas informações para o campo da glicobiologia, mais especificamente na N-glicosilação de proteínas como uma ferramenta multipropósito. Dentre as linhas pesquisadas, surgem as seguintes perspectivas:

1. Realizar investigações que contemplem o processo de ligação e de saída dos substratos da PglB, utilizando-se métodos que permitam também o cálculo da afinidade dos ligantes, como a metadinâmica e o *pulling* seguido de *umbrella sampling*. A compreensão das interações que regem essa ligação possibilitaria a introdução de novos substratos e uma manipulação mais fina do processo;
2. Testar a maleabilidade das enzimas quanto à troca entre unidades estruturais diferentes e os efeitos causados por essas alterações;
3. Construir uma segunda geração de PglBs quiméricas que possuam termoestabilidade com potencial biotecnológico para que possam utilizar sua promiscuidade com alta rentabilidade;
4. Empregar cálculos de QM para descrever o mecanismo catalítico da PglB, tornando palpável a modificação racional do sítio catalítico para que a enzima catalise eficientemente substratos distintos (em sinergia com o item 1).

Referências

- 1 VARKI, A. et al. *Essentials of Glycobiology*. [S.l.]: Cold Spring Harbor Laboratory Press, 2009. ISBN 9780879697709. Citado na página 16,
- 2 RADEMACHER T. W.; PAREKH, R. B. D. R. A. Glycobiology. *Ann. Rev. Biochem*, v. 57, p. 785–838, 1988. Citado na página 16,
- 3 NOTHAFT, H.; SZYMANSKI, C. M. Bacterial protein n-glycosylation: New perspectives and applications. *Journal of Biological Chemistry*, v. 288, n. 10, p. 6912–6920, 2013. Citado 3 vezes nas páginas 16, 18 e 22,
- 4 VARKI, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, v. 3, n. 2, p. 97–130, 1993. Citado na página 17,
- 5 SPIRO, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, v. 12, n. 4, p. 43R–56R, 2002. Citado na página 17,
- 6 APWEILER, R. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1473, n. 1, p. 4–8, 1999. Citado 2 vezes nas páginas 17 e 18,
- 7 NEUBERGER, A. Carbohydrates in protein: The carbohydrate component of crystalline egg albumin. *The Biochemical journal*, v. 32, n. 9, p. 1435–51, 1938. Citado na página 17,
- 8 SHENTAL-BECHOR, D.; LEVY, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. USA*, v. 105, n. 24, p. 8256–8261, 2008. Citado na página 17,
- 9 SCHACHTER, H.; FREEZE, H. H. Glycosylation diseases: Quo vadis? *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, v. 1792, n. 9, p. 925–930, 2009. Citado na página 17,
- 10 JAEKEN, J. Congenital disorders of glycosylation (CDG): it's (nearly) all in it! *Journal of Inherited Metabolic Disease*, Springer Netherlands, v. 34, n. 4, p. 853–858, 2011. Citado na página 17,
- 11 STEPPER, J. et al. Cysteine S -glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Letters*, v. 585, n. 4, p. 645–650, 2011. Citado na página 17,
- 12 POL-FACHIN, L.; VERLI, H. Assessment of Glycoproteins Dynamics from Computer Simulations. *Mini-Reviews in Organic Chemistry*, v. 8, p. 229–238, 2011. Citado na página 17,

- 13 NOTHAFT, H. et al. *Campylobacter jejuni* free oligosaccharides: Function and fate. *Virulence*, n. 6, p. 546–550. Citado 2 vezes nas páginas 18 e 22,
- 14 CALO, D.; KAMINSKI, L.; EICHLER, J. *Protein glycosylation in Archaea: Sweet and extreme*. 2010. Citado na página 18,
- 15 JOHANSEN, P. G.; MARSHALL, R. D.; NEUBERGER, A. Carbohydrates in protein. 3 The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *The Biochemical journal*, v. 78, n. 3, p. 518–27, 1961. Citado na página 18,
- 16 MESCHER, M. F.; STROMINGER, J. L. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *The Journal of Biological Chemistry*, v. 251, n. 7, p. 2005–14, 1976. Citado na página 18,
- 17 WIELAND, F.; PAUL, G.; SUMPER, M. Halobacterial flagellins are sulfated glycoproteins. *The Journal of biological chemistry*, v. 260, n. 28, p. 15180–5, 1985. Citado na página 18,
- 18 MULLAKHANBHAI, M. F.; LARSEN, H. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Archives of Microbiology*, v. 104, n. 3, p. 207–14, 1975. Citado na página 18,
- 19 SZYMANSKI, C. M. et al. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology*, v. 32, n. 5, p. 1022–1030, 1999. Citado na página 18,
- 20 YOUNG, N. M. et al. Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, *Campylobacter jejuni*. *Journal of Biological Chemistry*, v. 277, n. 45, p. 42530–42539, 2002. Citado 2 vezes nas páginas 18 e 23,
- 21 LEHLE, L.; STRAHL, S.; TANNER, W. Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angewandte Chemie International Edition*, v. 45, n. 41, p. 6802–6818, 2006. ISSN 14337851. Citado na página 18,
- 22 LARKIN, A.; IMPERIALI, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry*, v. 50, n. 21, p. 4411–4426, 2011. Citado 3 vezes nas páginas 18, 20 e 21,
- 23 HELENIUS, J. et al. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature*, Nature Publishing Group, v. 415, n. 6870, p. 447–450, 2002. Citado na página 20,
- 24 JELK, J. et al. Glycoprotein biosynthesis in a eukaryote lacking the membrane protein Rft1. *The Journal of Biological Chemistry*, American Society for Biochemistry and Molecular Biology, v. 288, n. 28, p. 20616–23, 2013. Citado na página 20,
- 25 KELLY, J. et al. Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *Journal of Bacteriology*, v. 188, n. 7, p. 2427–2434, 2006. Citado na página 20,

- 26 KAMINSKI, L. et al. AglR is required for addition of the final mannose residue of the N-linked glycan decorating the *Haloflex volcanii* S-layer glycoprotein. *Biochimica et biophysica acta*, v. 1820, n. 10, p. 1664–70, 2012. Citado na página 20,
- 27 DEMPSKI, R. E.; IMPERIALI, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Current Opinion in Chemical Biology*, v. 6, n. 6, p. 844–850, 2002. Citado na página 22,
- 28 KELLEHER, D. J. et al. Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Molecular Cell*, v. 12, n. 1, p. 101–111, 2003. Citado na página 22,
- 29 YURIST-DOUTSCH, S. et al. *aglF*, *aglG* and *aglI*, novel members of a gene island involved in the n-glycosylation of the *Haloflex volcanii* s-layer glycoprotein. *Molecular Microbiology*, v. 69, n. 5, p. 1234–1245, 2008. Citado 2 vezes nas páginas 22 e 24,
- 30 WACKER, M. et al. N-Linked Glycosylation in *Campylobacter jejuni* and Its Functional Transfer into *E. coli*. *Science*, v. 298, n. 5599, p. 1790–1793, 2002. Citado 2 vezes nas páginas 22 e 37,
- 31 NASAB, F. P. et al. All in One: *Leishmania major* STT3 Proteins Substitute for the Whole Oligosaccharyltransferase Complex in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell*, v. 19, p. 3758–3768, 2008. Citado na página 22,
- 32 MATSUMOTO, S.; NYIRENDA, J.; KOHDA, D. Structural Biology of Oligosaccharyltransferase (OST). In: *Glycoscience: Biology and Medicine*. Tokyo: Springer Japan, 2015. p. 437–445. Citado 3 vezes nas páginas 22, 27 e 126,
- 33 HERSCOVICS, A. Processing glycosidases of *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1426, n. 2, p. 275–285, 1999. Citado na página 22,
- 34 ROTH, J. Protein N-Glycosylation along the Secretory Pathway: Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chem. Rev.*, American Chemical Society, v. 102, n. 2, p. 285–304, 2002. Citado na página 22,
- 35 IELMINI, M. V.; FELDMAN, M. F. *Desulfovibrio desulfuricans* PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements†. *Glycobiology*, Oxford University Press, v. 21, n. 6, p. 734–742, 2011. Citado na página 22,
- 36 GRASS, S. et al. The *Haemophilus influenzae* HMW1C Protein Is a Glycosyltransferase That Transfers Hexose Residues to Asparagine Sites in the HMW1 Adhesin. *PLoS Pathogens*, Public Library of Science, v. 6, n. 5, p. e1000919, 2010. Citado na página 22,
- 37 SCHOENHOFEN, I. C. et al. Functional characterization of dehydratase/amino-transferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *The Journal of Biological Chemistry*, v. 281, n. 2, p. 723–32, 2006. Citado na página 23,

- 38 OLIVIER, N. B. et al. In vitro biosynthesis of UDP-N,N'-diacetylglucosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry*, v. 45, n. 45, p. 13659–13669, 2006. Citado na página 23,
- 39 GLOVER, K. J. et al. Direct biochemical evidence for the utilization of UDP-glucosamine by PglC, an essential glucosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry*, American Chemical Society, v. 45, n. 16, p. 5343–5350, 2006. Citado na página 23,
- 40 WEERAPANA, E.; IMPERIALI, B. Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology*, v. 16, n. 6, p. 91–101, 2006. Citado na página 23,
- 41 GLOVER, K. J. et al. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chemistry and Biology*, v. 12, n. 12, p. 1311–1316, 2005. Citado na página 23,
- 42 TROUTMAN, J. M.; IMPERIALI, B. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glucosyl transfer reactions. *Biochemistry*, v. 48, n. 12, p. 2807–2816, 2009. Citado na página 23,
- 43 KOWARIK, M. et al. Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO journal*, v. 25, n. 9, p. 1957–1966, 2006. Citado na página 23,
- 44 SCHWARZ, F. et al. Relaxed acceptor site specificity of bacterial oligosaccharyl-transferase in vivo. *Glycobiology*, v. 21, n. 1, p. 45–54, 2011. Citado 2 vezes nas páginas 23 e 24,
- 45 SILVERMAN, J. M.; IMPERIALI, B. Bacterial N-glycosylation efficiency is dependent on the structural context of target sequons. *Journal of Biological Chemistry*, 2016. ISSN 1083351X. Citado na página 23,
- 46 EICHLER, J. Extreme sweetness: protein glycosylation in archaea. *Nature Reviews Microbiology*, Nature Publishing Group, v. 11, n. 3, p. 151–156, 2013. Citado na página 24,
- 47 KAMINSKI, L. et al. Phylogenetic- and genome-derived insight into the evolution of N-glycosylation in Archaea. *Molecular Phylogenetics and Evolution*, Elsevier Inc., n. 2, p. 327–339. ISSN 10557903. Citado 2 vezes nas páginas 24 e 31,
- 48 GODIN, N.; EICHLER, J. Aglgenes, a curated and searchable database of archaeal N-glycosylation pathway components. *Database*, v. 2014, 2014. Citado na página 24,
- 49 EICHLER, J.; JARRELL, K.; ALBERS, S. A proposal for the naming of N-glycosylation pathway components in Archaea. *Glycobiology*, v. 23, n. 6, p. 620–621, 2013. Citado na página 24,
- 50 JARRELL, K. F. et al. N-Linked Glycosylation in Archaea: a Structural, Functional, and Genetic Analysis. *Microbiology and Molecular Biology Reviews*, 2014. Citado 2 vezes nas páginas 24 e 25,

- 51 SUMPER, M. et al. Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *Journal of Bacteriology*, v. 172, n. 12, p. 7111–7118, 1990. Citado na página 24,
- 52 EICHLER, J.; MAUPIN-FURLOW, J. Post-translation modification in Archaea: Lessons from *Haloferax volcanii* and other haloarchaea. *FEMS Microbiology Reviews*, v. 37, n. 4, p. 583–606, 2013. Citado na página 24,
- 53 KAMINSKI, L. et al. Add salt, add sugar: N-glycosylation in *Haloferax volcanii*. *Biochemical Society Transactions*, v. 41, n. 1, p. 432–5, 2013. Citado na página 24,
- 54 KAMINSKI, L. et al. AglJ adds the first sugar of the N-linked pentasaccharide decorating the *Haloferax volcanii* S-layer glycoprotein. *Journal of Bacteriology*, v. 192, n. 21, p. 5572–5579, 2010. Citado na página 24,
- 55 YURIST-DOUTSCH, S. et al. N-glycosylation in Archaea: On the coordinated actions of *Haloferax volcanii* AglF and AglM. *Molecular Microbiology*, v. 75, n. 4, p. 1047–1058, 2010. Citado na página 24,
- 56 MAGIDOVICH, H. et al. AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*. *Molecular Microbiology*, v. 76, n. 1, p. 190–199, 2010. Citado na página 24,
- 57 ARBIV, A. et al. AglQ Is a Novel Component of the *Haloferax volcanii* N-Glycosylation Pathway. *PLoS ONE*, v. 8, n. 11, p. e81782, 2013. Citado na página 24,
- 58 KAMINSKI, L. et al. Two distinct N-glycosylation pathways process the *Haloferax volcanii* S-layer glycoprotein upon changes in environmental salinity. *mBio*, 2013. Citado na página 26,
- 59 MEYER, B. H.; ALBERS, S. V. AglB, catalyzing the oligosaccharyl transferase step of the archaeal N-glycosylation process, is essential in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *MicrobiologyOpen*, 2014. Citado na página 26,
- 60 MEYER, B. H.; ALBERS, S.-V. Hot and sweet: protein glycosylation in Crenarchaeota. *Biochemical Society Transactions*, v. 41, n. 1, 2013. Citado na página 26,
- 61 LIZAK, C. et al. Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nature Communications*, 2013. Citado 8 vezes nas páginas 27, 33, 34, 35, 114, 125, 131 e 136,
- 62 IGURA, M. et al. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *The EMBO journal*, v. 27, n. 1, p. 234–43, 2008. Citado 3 vezes nas páginas 27, 29 e 126,
- 63 MAITA, N. et al. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *Journal of Biological Chemistry*, v. 285, n. 7, p. 4941–4950, 2010. Citado 3 vezes nas páginas 27, 29 e 126,

- 64 LIZAK, C. et al. X-ray structure of a bacterial oligosaccharyltransferase. *Nature*, v. 474, n. 7351, p. 350–355, 2011. Citado 10 vezes nas páginas 27, 28, 29, 32, 34, 35, 50, 114, 125 e 126,
- 65 MATSUMOTO, S. et al. Crystal structure of the c-terminal globular domain of oligosaccharyltransferase from archaeoglobus fulgidus at 1.75 Å resolution. *Biochemistry*, v. 51, n. 20, p. 4157–4166, 2012. Citado 5 vezes nas páginas 28, 29, 126, 127 e 131,
- 66 MATSUMOTO, S.; SHIMADA, A.; KOHDA, D. Crystal structure of the C-terminal globular domain of the third paralog of the Archaeoglobus fulgidus oligosaccharyltransferases. *BMC Structural Biology*, v. 13, n. 1, p. 11, 2013. Citado 7 vezes nas páginas 28, 29, 33, 34, 51, 133 e 134,
- 67 MORAES, I. *The Next Generation in Membrane Protein Structure Determination*. [S.l.: s.n.], 2016. v. 922. Citado na página 29,
- 68 ZUBKOV, S.; LENNARZ, W. J.; MOHANTY, S. Structural basis for the function of a minimembrane protein subunit of yeast oligosaccharyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences, v. 101, n. 6056, p. 3821–3826, 2004. Citado na página 29,
- 69 SCHULZ, B. L. et al. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences, v. 106, n. 27, p. 11061–6, 2009. Citado na página 29,
- 70 NYIRENDA, J. et al. Crystallographic and NMR evidence for flexibility in oligosaccharyltransferases and its catalytic significance. *Structure*, Elsevier Ltd, v. 21, n. 1, p. 32–41, 2013. Citado 2 vezes nas páginas 29 e 133,
- 71 HUANG, C.; BHASKARAN, R.; MOHANTY, S. Eukaryotic N-glycosylation occurs via the membrane-anchored C-terminal domain of the Stt3p subunit of oligosaccharyltransferase. *Journal of Biological Chemistry*, v. 287, n. 39, p. 32450–32458, 2012. Citado 2 vezes nas páginas 29 e 126,
- 72 YAN, Q.; LENNARZ, W. J. Studies on the function of oligosaccharyl transferase subunits: A glycosylatable photoprobe binds to the luminal domain of Ost1p. *Proceedings of the National Academy of Sciences*, v. 99, n. 25, p. 15994–15999, 2002. Citado na página 31,
- 73 JAFFEE, M. B.; IMPERIALI, B. Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound N-linked oligosaccharyl transferases. *Biochemistry*, v. 50, n. 35, p. 7557–7567, 2011. Citado na página 32,
- 74 GERBER, S. et al. Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *Journal of Biological Chemistry*, 2013. Citado 2 vezes nas páginas 32 e 33,
- 75 OLLIS, A. A. et al. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nature Chemical Biology*, 2014. Citado na página 32,

- 76 OLLIS, A. A. et al. Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Scientific Reports*, 2015. Citado na página 32,
- 77 MATSUMOTO, S. et al. Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 44, p. 17868–73, 2013. Citado 2 vezes nas páginas 32 e 36,
- 78 MATSUMOTO, S. et al. Tethering an N-glycosylation sequon-containing peptide creates a catalytically competent oligosaccharyltransferase complex. *Biochemistry*, 2017. Citado 2 vezes nas páginas 32 e 51,
- 79 LAIRSON, L. et al. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, v. 77, n. 1, p. 521–555, 2008. Citado 3 vezes nas páginas 33, 34 e 36,
- 80 WIGGINS, C. A. R.; MUNRO, S. Activity of the yeast MNN1 -1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proceedings of the National Academy of Sciences*, v. 95, n. 14, p. 7945–7950, 1998. Citado 2 vezes nas páginas 34 e 116,
- 81 LIZAK, C. et al. A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *Journal of Biological Chemistry*, v. 289, n. 2, p. 735–746, 2014. Citado 3 vezes nas páginas 36, 113 e 133,
- 82 BAKER, J. L.; ÇELIK, E.; DELISA, M. P. Expanding the glycoengineering toolbox: The rise of bacterial N-linked protein glycosylation. *Trends in Biotechnology*, v. 31, n. 5, p. 313–323, 2013. Citado 3 vezes nas páginas 36, 37 e 39,
- 83 KEYS, T. G.; AEBI, M. Engineering protein glycosylation in prokaryotes Q1 Q5. *Current Opinion in Systems Biology*, p. 1–9, 2017. Citado na página 36,
- 84 SETHURAMAN, N.; STADHEIM, T. A. Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology*, v. 17, n. 4, p. 341–346, 2006. Citado na página 36,
- 85 TERRA, V. S. et al. *Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design*. 2012. Citado na página 37,
- 86 FELDMAN, M. F. et al. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 102, n. 8, p. 3016–3021, 2005. Citado na página 37,
- 87 IHSEN, J. et al. Production of glycoprotein vaccines in *Escherichia coli*. *Microbial Cell Factories*, v. 9, n. 1, p. 61, 2010. Citado na página 37,
- 88 IHSEN, J. et al. Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biology*, v. 5, n. 4, p. 140227, 2015. Citado na página 37,

- 89 GARCIA-QUINTANILLA, F. et al. Production of a recombinant vaccine candidate against *Burkholderia pseudomallei* exploiting the bacterial N-glycosylation machinery. *Frontiers in Microbiology*, 2014. Citado na página 37,
- 90 PRICE, N. L. et al. Glycoengineered Outer Membrane Vesicles: A Novel Platform for Bacterial Vaccines. *Scientific Reports*, 2016. Citado 2 vezes nas páginas 37 e 39,
- 91 SRICHAISUPAKIT, A. et al. Production of initial-stage eukaryotic N-glycan and its protein glycosylation in *Escherichia coli*. *Journal of Bioscience and Bioengineering*, 2015. ISSN 13474421. Citado na página 37,
- 92 HUG, I.; FELDMAN, M. F. *Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria*. 2011. Citado na página 37,
- 93 VALDERRAMA-RINCON, J. D. et al. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nature Chemical Biology*, v. 8, n. 5, p. 434–436, 2012. Citado na página 37,
- 94 FOCARETA, A. et al. A Recombinant Probiotic for Treatment and Prevention of Cholera. *Gastroenterology*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, v. 130, n. 6, p. 1688–1695, 2006. Citado na página 37,
- 95 MALLY, M. et al. Glycoengineering of host mimicking type-2 LacNAc polymers and Lewis X antigens on bacterial cell surfaces. *Molecular Microbiology*, v. 87, n. 1, p. 112–131, 2013. Citado na página 37,
- 96 YAVUZ, E. et al. Glycomimicry: Display of fucosylation on the lipo-oligosaccharide of recombinant *Escherichia coli* K12. *Glycoconjugate Journal*, v. 28, n. 1, p. 39–47, 2011. Citado na página 37,
- 97 WACKER, M. et al. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proceedings of the National Academy of Sciences*, 2006. Citado na página 37,
- 98 CHEN, M. M. et al. Polyisoprenol Specificity in the *Campylobacter jejuni* N-Linked Glycosylation Pathway. *Biochemistry*, v. 46, n. 50, p. 14342–14348, 2007. Citado na página 37,
- 99 CURRY, S. Structural Biology: A Century-long Journey into an Unseen World. *Interdisciplinary Science Reviews*, v. 40, n. 3, p. 308–328, 2015. Citado na página 38,
- 100 HOSPITAL ADAM; GOÑI, J. O. M. G. J. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, v. 8, p. 37–47, 2015. Citado 3 vezes nas páginas 38, 42 e 45,
- 101 FISCHER, H. et al. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography*, International Union of Crystallography, n. 1, p. 101–109, feb. ISSN 00218898. Citado na página 38,

- 102 GREENFIELD, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, v. 1, n. 6, p. 2876–2890, 2007. Citado na página 38,
- 103 KRAUSS, I. R. et al. An Overview of Biological Macromolecule Crystallization. *International Journal of Molecular Sciences*, Multidisciplinary Digital Publishing Institute, v. 14, n. 6, p. 11643–11691, 2013. Citado na página 38,
- 104 WLODAWER, A. et al. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *The FEBS journal*, v. 280, n. 22, p. 5705–36, 2013. Citado na página 38,
- 105 MALUF, F. V. et al. Cristalografia. In: *Bioinformática: da Biologia à Flexibilidade Moleculares*. 1ª. ed. [S.l.: s.n.], 2014. cap. 13, p. 282. ISBN 978-85-69288-00-8. Citado na página 38,
- 106 ZIAREK, J. J.; BAPTISTA, D.; WAGNER, G. Recent developments in solution nuclear magnetic resonance (NMR)-based molecular biology. *Journal of Molecular Medicine*, Journal of Molecular Medicine, p. 1–8, 2017. Citado na página 40,
- 107 KWAN, E. E.; HUANG, S. G. Structural Elucidation with NMR Spectroscopy: Practical Strategies for Organic Chemists. *European Journal of Organic Chemistry*, v. 2008, n. 16, p. 2671–2688, 2008. Citado na página 40,
- 108 ALMEIDA, M. d. S. Ressonância Magnética Nuclear. In: *Bioinformática: da Biologia à Flexibilidade Moleculares*. 1ª. ed. [S.l.: s.n.], 2014. cap. 12, p. 282. ISBN 978-85-69288-00-8. Citado na página 40,
- 109 FIAUX, J. et al. NMR analysis of a 900K GroEL–GroES complex. *Nature*, Nature Publishing Group, v. 418, n. 6894, p. 207–211, 2002. Citado na página 40,
- 110 SPRANGERS, R.; KAY, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*, v. 445, n. 7128, p. 618–622, 2007. Citado na página 40,
- 111 HIGGINS, M. K.; LEA, S. M. On the state of crystallography at the dawn of the electron microscopy revolution. *Current Opinion in Structural Biology*, v. 46, p. 95–101, 2017. Citado na página 40,
- 112 CAMPBELL, M. G. et al. 2.8 Å resolution reconstruction of the thermoplasma acidophilum 20 s proteasome using cryo-electron microscopy. *eLife*, v. 2015, n. 4, p. 1–22, 2015. Citado na página 40,
- 113 FISCHER, N. M. et al. Properties of Organic Liquids when Simulated with Long-Range Lennard-Jones Interactions. *Journal of Chemical Theory and Computation*, American Chemical Society, v. 11, n. 7, p. 2938–2944, 2015. Citado na página 40,
- 114 VIJAY-KUMAR, S.; BUGG, C.; COOK, W. Structure of ubiquitin refined at 1.8 Å resolution. *J.Mol.Biol.*, v. 194, p. 531–544, 1987. Citado na página 41,

- 115 YANG, Y. et al. Solution structure of proinsulin: Connecting domain flexibility and prohormone processing. *Journal of Biological Chemistry*, American Society for Biochemistry and Molecular Biology, v. 285, n. 11, p. 7847–7851, 2010. Citado na página 41,
- 116 IACOVACHE, I. et al. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nature Communications*, v. 7, p. 12062, 2016. Citado na página 41,
- 117 BOHNE-LANG, A.; Von der Lieth, C. W. GlyProt: In silico glycosylation of proteins. *Nucleic Acids Research*, v. 33, n. SUPPL. 2, p. W214–W219, 2005. Citado na página 42,
- 118 PETRESCU, A. J.; WORMALD, M. R.; DWEK, R. A. *Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding*. 2006. 600–607 p. Citado na página 42,
- 119 FACHIN, P. *Glicobiologia estrutural da modulação da cascata de coagulação sanguínea por heparina*. 171 p. Tese (Doutorado), 2013. Citado na página 42,
- 120 WOODS, R. J. *Computational carbohydrate chemistry: What theoretical methods can tell us*. 1998. 209–216 p. Citado na página 42,
- 121 PETRESCU, A. J. et al. Statistical analysis of the protein environment of N-glycosylation sites: Implications for occupancy, structure, and folding. *Glycobiology*, v. 14, n. 2, p. 103–114, 2004. Citado na página 42,
- 122 LÜTTEKE, T.; FRANK, M.; Von Der Lieth, C. W. Data mining the protein data bank: Automatic detection and assignment of carbohydrate structures. In: *Carbohydrate Research*. [S.l.: s.n.], 2004. v. 339, n. 5, p. 1015–1020. Citado na página 42,
- 123 CRISPIN, M.; STUART, D. I.; JONES, E. Y. Building meaningful models of glycoproteins. *Nature structural & Molecular Biology*, Nature Publishing Group, v. 14, n. 5, p. 354, 2007. Citado na página 42,
- 124 GUNSTEREN, W. F. van et al. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie (International ed. in English)*, v. 45, n. 25, p. 4064–92, 2006. Citado na página 42,
- 125 TOZZINI, V. Multiscale Modeling of Proteins. *Accounts of Chemical Research*, American Chemical Society, v. 43, n. 2, p. 220–230, 2010. Citado na página 42,
- 126 DROR, R. O. et al. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, p. 429–52. ISSN 1936-1238. Citado na página 42,
- 127 POL-FACHIN, L.; FERNANDES, C. L.; VERLI, H. GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydrate Research*, v. 344, n. 4, p. 491–500, 2009. Citado na página 42,

- 128 FADDA, E.; WOODS, R. J. *Molecular simulations of carbohydrates and protein-carbohydrate interactions: Motivation, issues and prospects*. 2010. 596–609 p. Citado na página 42,
- 129 FERNANDES, C. L. et al. GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins. *Carbohydrate Research*, v. 345, n. 5, p. 663–671, 2010. Citado 2 vezes nas páginas 42 e 125,
- 130 PEDEBOS, C.; POL-FACHIN, L.; VERLI, H. Unrestrained conformational characterization of *Stenocereus eruca* saponins in aqueous and nonaqueous solvents. *Journal of Natural Products*, v. 75, n. 6, p. 1196–1200, 2012. Citado na página 42,
- 131 HRICOVINI, M. *Carbohydrate Molecular Dynamics: Oligosaccharides, Polysaccharides, and Glycoconjugates*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. 222–228 p. Citado na página 42,
- 132 ARANTES, P. R. et al. Conformational characterization of ipomotaosides and their recognition by COX-1 and 2. *Molecules*, Multidisciplinary Digital Publishing Institute, v. 19, n. 4, p. 5421–5433, 2014. Citado na página 42,
- 133 MALLAJOSYULA, S. S.; JO, S.; IM, W. Glycoinformatics. *Methods in molecular biology (Clifton, N.J.)*, v. 1273, p. 407–429, 2015. Citado na página 42,
- 134 WARSHEL1976. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, v. 103, n. 2, p. 227–249, 1976. Citado na página 42,
- 135 MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. *Nature*, v. 267, n. 5612, p. 585–590, 1977. Citado na página 42,
- 136 TINOCO, I.; WEN, J.-D. Simulation and analysis of single-ribosome translation. *Physical biology*, v. 6, n. 2, p. 025006, 2009. Citado na página 42,
- 137 HUBER, R. G. et al. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Progress in Biophysics and Molecular Biology*, Elsevier Ltd. ISSN 00796107. Citado na página 42,
- 138 BEST, R. B. *Computational and theoretical advances in studies of intrinsically disordered proteins*. 2017. 147–154 p. Citado na página 42,
- 139 STANSFELD, P. J. Computational studies of membrane proteins: from sequence to structure to simulation. *Current Opinion in Structural Biology*, Elsevier Ltd, v. 45, p. 133–141, 2017. Citado na página 42,
- 140 LEACH, A. R. *Molecular modelling : principles and applications*. [S.l.]: Prentice Hall, 2001. 744 p. ISBN 0582382106. Citado 2 vezes nas páginas 42 e 43,
- 141 CORNELL, W. D. et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, American Chemical Society, v. 117, n. 19, p. 5179–5197, 1995. Citado na página 43,

- 142 MACKERELL, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, American Chemical Society, v. 102, n. 18, p. 3586–3616, 1998. Citado na página 44,
- 143 OOSTENBRINK, C. et al. Validation of the 53A6 GROMOS force field. *European Biophysics Journal*, Springer-Verlag, v. 34, n. 4, p. 273–284, 2005. Citado na página 44,
- 144 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, American Chemical Society, v. 118, n. 45, p. 11225–11236, 1996. Citado na página 44,
- 145 NG, H. W.; LAUGHTON, C. A.; DOUGHTY, S. W. Molecular dynamics simulations of the adenosine A2a receptor: Structural stability, sampling, and convergence. *Journal of Chemical Information and Modeling*, 2013. Citado 2 vezes nas páginas 45 e 60,
- 146 HESS, B. Convergence of sampling in protein simulations. *Physical Review E*, v. 65, n. 3, p. 031910, 2002. Citado na página 45,
- 147 FARALDO-GÓMEZ, J. D. et al. Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins: Structure, Function and Genetics*, v. 57, n. 4, p. 783–791, 2004. Citado na página 45,
- 148 GROSSFIELD, A.; FELLER, S. E.; PITMAN, M. C. Convergence of molecular dynamics simulations of membrane proteins. *Proteins*, n. 1, p. 31–40, jan. ISSN 1097-0134. Citado na página 45,
- 149 ROMO, T. D.; GROSSFIELD, A. Block covariance overlap method and convergence in molecular dynamics simulation. *Journal of Chemical Theory and Computation*, American Chemical Society, v. 7, n. 8, p. 2464–2472, 2011. Citado na página 45,
- 150 LYMAN, E.; ZUCKERMAN, D. M. On the structural convergence of biomolecular simulations by determination of the effective sample size. *Journal of Physical Chemistry B*, v. 111, n. 44, p. 12876–12882, 2007. Citado na página 45,
- 151 KRISHNAN, R.; WALTON, E. B.; Van Vliet, K. J. Characterizing rare-event property distributions via replicate molecular dynamics simulations of proteins. *Journal of Molecular Modeling*, v. 15, n. 11, p. 1383–1389, 2009. Citado na página 45,
- 152 BARDUCCI, A.; BONOMI, M.; PARRINELLO, M. *Metadynamics*. 2011. 826–843 p. Citado 2 vezes nas páginas 45 e 56,
- 153 SUGITA, Y.; OKAMOTO, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, v. 314, n. 1-2, p. 141–151, 1999. Citado na página 45,
- 154 NOÉ, F. et al. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 45, p. 19011–6, 2009. Citado na página 45,

- 155 NAGANATHAN, A. N.; OROZCO, M. The native ensemble and folding of a protein Molten-globule: Functional consequence of downhill folding. *Journal of the American Chemical Society*, v. 133, n. 31, p. 12154–12161, 2011. Citado na página 45,
- 156 CANDOTTI, M. et al. Toward an atomistic description of the urea-denatured state of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 15, p. 5933–8, 2013. Citado na página 45,
- 157 FENWICK, R. B.; ESTEBAN-MARTÍN, S.; SALVATELLA, X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *European Biophysics Journal*, v. 40, n. 12, p. 1339–1355, 2011. Citado na página 45,
- 158 KERN, N. R. et al. Lipid-linked oligosaccharides in membranes sample conformations that facilitate binding to oligosaccharyltransferase. *Biophysical Journal*, v. 107, n. 8, p. 1885–1895, 2014. Citado 3 vezes nas páginas 45, 52 e 110,
- 159 PEDEBOS, C. et al. Glicobiologia Estrutural e Evolução Molecular da PglB de *Campylobacter lari*. 2013. Citado 3 vezes nas páginas 46, 62 e 125,
- 160 PEDEBOS, C. et al. *In silico Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination*. [S.l.: s.n.], 2015. v. 25. 1183–1195 p. Citado 3 vezes nas páginas 46, 62 e 125,
- 161 LEE, H. S.; IM, W. Transmembrane motions of PglB induced by LLO are coupled with EL5 loop conformational changes necessary for OST activity. *Glycobiology*, v. 277, p. 1–9, 2017. Citado 4 vezes nas páginas 46, 109, 113 e 133,
- 162 KO, J. et al. The FALC-Loop web server for protein loop modeling. *Nucleic Acids Research*, v. 39, n. SUPPL. 2, p. W210–W214, 2011. Citado na página 51,
- 163 LEE, J. et al. Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins: Structure, Function and Bioinformatics*, v. 78, n. 16, p. 3428–3436, 2010. Citado na página 51,
- 164 LASKOWSKI, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Research*, v. 29, n. 1, p. 221–2, 2001. Citado na página 51,
- 165 SLYNKO, V. et al. NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. *Journal of the American Chemical Society*, v. 131, n. 3, p. 1274–1281, 2009. Citado na página 51,
- 166 WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. In: *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016. v. 54, p. 5.6.1–5.6.37. Citado na página 51,
- 167 SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, v. 7, n. 1, p. 539, 2011. Citado 2 vezes nas páginas 51 e 59,
- 168 MORRIS, G. M. et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, v. 30, n. 16, p. 2785–2791, 2009. Citado na página 52,

- 169 SCHNEIDMAN-DUHOVNY, D. et al. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, v. 33, n. SUPPL. 2, p. 363–367, 2005. Citado na página 52,
- 170 KANDT, C.; ASH, W. L.; Peter Tieleman, D. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, v. 41, n. 4, p. 475–488, 2007. Citado na página 53,
- 171 SCHMIDT, T. H.; KANDT, C. LAMBADA and InflateGRO2: Efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations. *Journal of Chemical Information and Modeling*, v. 52, n. 10, p. 2657–2669, 2012. Citado na página 53,
- 172 PRONK, S. et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, v. 29, n. 7, p. 845–854, 2013. Citado na página 53,
- 173 OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry*, v. 25, n. 13, p. 1656–76, 2004. Citado na página 54,
- 174 POL-FACHIN, L. et al. GROMOS 53A6 GLYC , an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *Journal of Chemical Theory and Computation*, American Chemical Society, v. 8, n. 11, p. 4681–4690, 2012. Citado 2 vezes nas páginas 54 e 56,
- 175 POL-FACHIN, L.; VERLI, H.; LINS, R. D. Extension and validation of the GROMOS 53A6(GLYC) parameter set for glycoproteins. *Journal of computational chemistry*, v. 35, n. 29, p. 2087–95, 2014. Citado 2 vezes nas páginas 54 e 56,
- 176 DOMAŃSKI, J. et al. Lipidbook: A public repository for force-field parameters used in membrane simulations. *Journal of Membrane Biology*, v. 236, n. 3, p. 255–258, 2010. Citado na página 54,
- 177 PIGGOT, T. J.; HOLDBROOK, D. A.; KHALID, S. Electroporation of the E. coli and S. aureus membranes: Molecular dynamics simulations of complex bacterial membranes. *Journal of Physical Chemistry B*, 2011. Citado na página 54,
- 178 BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry*, v. 91, n. 24, p. 6269–6271, 1987. Citado na página 54,
- 179 DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, v. 98, n. 12, p. 10089, 1993. Citado 2 vezes nas páginas 54 e 55,
- 180 HESS, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation*, American Chemical Society, v. 4, n. 1, p. 116–122, 2008. Citado na página 54,

- 181 BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, v. 126, n. 1, p. 014101, 2007. Citado na página 55,
- 182 NOSÉ, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, Taylor & Francis Group, v. 52, n. 2, p. 255–268, 1984. Citado na página 55,
- 183 HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, American Physical Society, v. 31, n. 3, p. 1695–1697, 1985. Citado na página 55,
- 184 PARRINELLO, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, v. 52, n. 12, p. 7182, 1981. Citado na página 55,
- 185 NOSÉ, S.; KLEIN, M. L. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, Taylor & Francis Group, v. 50, n. 5, p. 1055–1076, 1983. Citado na página 55,
- 186 MILLER, W. G. et al. The complete genome sequence and analysis of the human pathogen *Campylobacter lari*. *Foodborne pathogens and disease*, v. 5, n. 4, p. 371–386, 2008. Citado na página 55,
- 187 KLENK, H.-P. et al. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Citado na página 55,
- 188 BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, AIP Publishing, v. 81, n. 8, p. 3684, 1984. Citado na página 55,
- 189 GUNSTEREN, W. F. van; BERENDSEN, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English*, Hüthig & Wepf Verlag, v. 29, n. 9, p. 992–1023, 1990. Citado na página 55,
- 190 CALEMAN, C. et al. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *Journal of chemical theory and computation*, n. 1, p. 61–74, jan. ISSN 1549-9626. Citado 2 vezes nas páginas 55 e 57,
- 191 BONOMI, M. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, Elsevier B.V., v. 180, n. 10, p. 1961–1972, 2009. Citado na página 56,
- 192 XU, J.; ZHANG, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, v. 26, n. 7, p. 889–895, 2010. Citado na página 56,

- 193 SCHMIDT, M. W. et al. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, v. 14, n. 11, p. 1347–1363, 1993. Citado 2 vezes nas páginas 56 e 57,
- 194 SCHMID, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European biophysics journal : EBJ*, v. 40, n. 7, p. 843–56, 2011. Citado na página 56,
- 195 SCHAFTENAAR, G.; NOORDIK, J. H. Molden: A pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, Kluwer Academic Publishers, v. 14, n. 2, p. 123–134, 2000. Citado na página 57,
- 196 LINS, R. D. *Biomat - Rotational Profiler*. Citado na página 57,
- 197 SAYERS, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, v. 40, n. D1, p. D13–D25, 2012. Citado na página 57,
- 198 ALTSCHUL, S. F. et al. Basic local alignment search tool. *Journal of Molecular Biology*, v. 215, n. 3, p. 403–10, 1990. Citado na página 57,
- 199 FINN, R. D. et al. *Pfam: The protein families database*. [S.l.]: Oxford University Press, 2014. D222–30 p. Citado na página 59,
- 200 WHELAN, S.; GOLDMAN, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, v. 18, n. 5, p. 691–699, 2001. Citado na página 59,
- 201 DARRIBA, D.; TABOADA, G. L.; POSADA, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)*, Europe PMC Funders, v. 27, n. 8, p. 1164–1165, 2011. Citado na página 59,
- 202 KUMAR, S. et al. MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*, v. 28, n. 20, p. 2685–2686, 2012. Citado na página 59,
- 203 ABDI, H.; WILLIAMS, L. J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 2, n. 4, p. 433–459, 2010. Citado na página 60,
- 204 VERLI, H.; GUIMARÃES, J. A. Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydrate Research*, v. 339, n. 2, p. 281–290, 2004. Citado na página 125,
- 205 BECKER, C. F.; GUIMARÃES, J. A.; VERLI, H. Molecular dynamics and atomic charge calculations in the study of heparin 0conformation in aqueous solution. *Carbohydrate Research*, v. 340, n. 8, p. 1499–1507, 2005. Citado na página 125,
- 206 POL-FACHIN, L.; SERRATO, R. V.; VERLI, H. Solution conformation and dynamics of exopolysaccharides from Burkholderia species. *Carbohydrate Research*, v. 345, n. 13, p. 1922–1931, 2010. Citado na página 125,

- 207 POL-FACHIN, L. et al. Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Proteins: Structure, Function and Bioinformatics*, v. 79, n. 9, p. 2735–2745, 2011. Citado na página 125,
- 208 LIGABUE-BRAUN, R. et al. The Calcium Goes Meow: Effects of Ions and Glycosylation on Fel d 1, the Major Cat Allergen. *Plos One*, v. 10, n. Fig 1, p. e0132311, 2015. Citado na página 125,
- 209 MATSUMOTO, S.; SHIMADA, A.; KOHDA, D. Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Structural Biology*, v. 13, p. 1, 2013. Citado na página 126,
- 210 LOMBARD, J. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biology Direct*, v. 11, n. 1, p. 36, 2016. Citado 2 vezes nas páginas 127 e 129,
- 211 GOLDMAN, N. Maximum Likelihood Inference of Phylogenetic Trees, with Special Reference to a Poisson Process Model of DNA Substitution and to Parsimony Analyses. *Syst. Zool*, v. 39, n. 4, p. 345–361, 1990. Citado na página 127,
- 212 HOLDER, M.; LEWIS, P. O. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, v. 4, n. 4, p. 275–284, 2003. Citado na página 127,
- 213 KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, v. 10, n. 6, p. 845–858, 2015. Citado na página 128,
- 214 ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, v. 5, n. 4, p. 725–738, 2010. Citado na página 128,
- 215 LARA, P. et al. Refined topology model of the STT3/Stt3 protein subunit of the oligosaccharyl transferase complex. *Journal of Biological Chemistry*, p. jbc.M117.779421, 2017. Citado na página 130,
- 216 MARTÍ-RENOM, M. A. et al. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure*, v. 29, n. 1, p. 291–325, 2000. Citado na página 131,
- 217 RAMACHANDRAN, G.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, v. 7, n. 1, p. 95–99, 1963. Citado na página 131,

Anexos

ANEXO A – ABP1

Durante o Doutorado tive a oportunidade de realizar um trabalho em colaboração com a Dra. Cibele Tesser da Costa e o Prof. Arthur Fett-Neto (Laboratório de Fisiologia Vegetal) estudando a proteína ligadora de auxina 1 (auxin binding protein 1 - ABP1) por simulações de DM. Este trabalho está atualmente submetido à revista *Glycobiology*.

The role of Zn⁺², dimerization and N-glycosylation in the interaction of Auxin Binding Protein 1 (ABP1) with different auxins

Cibele Tesser da Costa, Conrado Pedebos, Hugo Verli, Arthur Germano Fett-Neto

Abstract

Auxin is critical for plant growth and development. The main natural auxin is indole-3-acetic acid (IAA), whereas 1-naphthalene acetic acid (NAA) is a synthetic form. Auxin Binding Protein 1 (ABP1) specifically binds auxins, presumably playing roles as receptor in non-transcriptional cell responses, although this role has been recently challenged by genetic analyses. ABP1 structure was previously established from maize at 1.9 Å resolution. To gain further insight on ABP1 structural biology, this study was carried out employing molecular dynamics simulations of the complete models of the oligomeric glycosylated proteins from maize and *Arabidopsis thaliana* with or without auxins. In maize, both Zn⁺² coordination and glycosylation promoted increased conformational stability and most of such stabilization effect was located on N- and C-terminal regions. The α -helix of C-terminal regions in ABP1 of both species unfolded during simulations, assuming a more extended structure in maize. In *Arabidopsis*, the helix appeared more stable, being preserved in most of the monomeric simulations and unfolding when the protein was in the dimeric form. In *Arabidopsis* ABP1 bound to IAA or NAA, glycosylation structures arranged around the protein, covering the putative site of entrance or egress of auxin. NAA bound protein folding was more similar to the crystal structure showing higher stability compared to IAA bound. The molecular structural differences of ABP1 found between the species and auxin types indicate that this auxin binding protein shows functional specificities in dicots and monocots, as well as in auxin type binding, which may reflect on its biological effects.

ANEXO B – DDB1 e DDB2

Outra colaboração que participei no período do Doutorado foi com o aluno MSc. Bruno César Feltes e o Prof. Diego Bonatto estudando (também por simulações de DM) os efeitos de mutações que ocorrem naturalmente no complexo DDB1-DDB2 causadoras da doença Xeroderma pigmentoso. Esse trabalho está atualmente em fase final de preparação para submissão à revista *Nucleic Acid Research*.

Dynamics of DDB2-DDB1 complex: structural origin for the role of mutations into Xeroderma Pigmentosum disease

Bruno César Feltes¹, Conrado Pedebos¹, Hugo Verli^{1*} and Diego Bonatto¹

Abstract: Xeroderma Pigmentosum (XP) is a disease caused by mutations in the nucleotide excision repair (NER) pathway, a molecular mechanism responsible for removing UV-induced DNA lesions. In this process, the DNA lesion recognition is performed by the DDB-Complex, composed by DDB2 (XPE) and DDB1. However, not much is known about how those mutations affect XP proteins structure and complex assembly. Thus, we searched for structural evidences associated to the role of three naturally occurring mutations found in DDB2 on XPE patients, R273H, K244Glu, and L350P, on the DDB-complex. Through a series of molecular dynamics simulations, DDB2 mutation promoted loss of flexibility, as well as loss of key movements of both domains, producing different a conformational behavior in comparison to the WT. In addition, a region comprising residues 354 to 371 was affected in the mutants. The DDB complex, containing the mutated forms of DDB2, showed three distinct behaviors: R273H displayed higher instability when complexed; L350P affected DDB1 protein-protein binding with DDB2, changing the motion all of the three, β -propellers of DDB1; and K244P, similar to the L350P mutant, loss all interacting residues when compared to the WT. Distinct domain motions were observed for DDB2 and DDB1 in each variant, explaining how the study mutations can impact on the DDB-Complex conformation, generating instability and impairing its proper function.

CURRICULUM VITÆ

MSc. Conrado Pedebos

PhD Student

Av. Bento Gonçalves, 9500
Porto Alegre, RS 91501-970
Campus do Vale - Building 43431
(51) 3308-7770 - conrado.pedebos@ufrgs.br

EDUCATION

Universidade Federal do Rio Grande do Sul, UFRGS, Brazil - *PhD in Progress*

2013 - 2017

Cellular and Molecular Biology

Advisor: Prof. Hugo Verli

Grantee: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brazil

Universidade Federal do Rio Grande do Sul, UFRGS, Brazil - *Master*

2012 - 2013

Master's Degree in Cellular and Molecular Biology

Title: Glicobiologia Estrutural e Evolução Molecular da PglB de *Campylobacter lari*

Advisor: Prof. Hugo Verli

Grantee: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brazil

Universidade Federal do Rio Grande do Sul, UFRGS, Brazil - *Bachelor*

2007 - 2011

Bachelor's Degree in Biomedicine

Title: Caracterização Conformacional de Saponinas

Advisor: Prof. Hugo Verli

Grantee: Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, FAPERGS, Brazil

SKILLS

Glycobiology - Macromolecule Chemistry - Structural Biology - Computational Biology - Phylogenetics - Computational Chemistry - Molecular Dynamics

RESEARCH PROJECTS

1. Computational Strategies for the Construction of a Carbohydrate Conformational and Modelling Database

2. Structure-activity Relationships and Biological Actions of G Protein-Coupled Receptors and their Ligands - CAPES/University of Nottingham - Drug Discovery (N°041/2014)

3. Characterization of the N-glycosylation Process and its Impact on the Function and Engineering of Proteins - SDumont LNCC/MCTI

4. Rational Development of New Candidates of Prototypes Acting as Anticoagulant Drugs

SCIENTIFIC PRODUCTION

Article Publications

1. **Pedebos, Conrado;** Arantes, Pablo R.; Giesel, Guilherme M.; Verli, Hugo. In silico Investigation of the PglB Active Site Reveals Transient Catalytic States and Octahedral Metal Ion Coordination. *Glycobiology*, v. 25 (11), p. 1183- 1195, 2015.
2. **Pedebos, Conrado;** Pol-Fachin, Laércio; Pons, Ramon; Teixeira, Cilaine; Verli, Hugo. Atomic Model and Micelle Dynamics of QS-21 Saponin. *Molecules*, v. 19, p. 3744-3760, 2014.
3. De Paula, Fernando T.; Frauches, Petrina Q.; **Pedebos, Conrado;** Berger, Markus; Gnoatto, Simone C. B.; Gosmann, Grace; Verli, Hugo; Guimaraes, Jorge A.; Graebin, Cedric S. Improving the Thrombin Inhibitory Activity of Glycyrrhizin, a Triterpenic Saponin, through a Molecular Simplification of the Carbohydrate Moiety. *Chemical Biology and Drug Design*, v. 82, p. 756-760, 2013.
4. **Pedebos, Conrado;** Pol-Fachin, Laércio; Verli, Hugo. Unrestrained Conformational Characterization of *Stenocereus eruca* Saponins in Aqueous and Nonaqueous Solvent. *Journal of Natural Products*, v. 75, p. 1196-1200, 2012.

Summaries published in proceedings of conferences

1. **Pedebos, C.;** Ligabue-Braun, R.; Arantes, P. R.; Verli, H. Evolution and Dynamics of the N-glycosylation Pathway through Oligosaccharyltransferases. XLV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2016, Natal / RN.
2. Arantes, P. R.; **Pedebos, C.;** Verli, H. Dynamics Of Lipid-Linked, Membrane Soaked, Oligosaccharides: Biological Precursors For N-Glycosylation In Eukarya And Prokarya. XLV Reunião Anual da

Sociedade Brasileira de Bioquímica e Biologia Molecular, 2016, Natal / RN.

3. **Pedebos, C.;** Verli, H. Insights on the Dynamics and Thermostability of an Archaeal Oligosaccharyltransferase. Programa de Resumos da XLIV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2015, Foz do Iguaçu / PR.
4. **Pedebos, C.;** Verli, H. Insights on the Dynamics and Thermostability of an Archaeal Oligosaccharyltransferase. Revista de Informática Teórica e Aplicada, 23 (1), 2016 - Conference Reports: Escola Gáucha de Bioinformática, 2015, Porto Alegre / RS.
5. Da Costa, C. T.; **Pedebos, C.;** Verli, H.; Fett-Neto, A. G. Effects of Zn⁺², N-Glycosylation and Dimerization on Auxin Binding Protein 1 (ABP1) Conformation and its Interaction with Auxin. Revista de Informática Teórica e Aplicada, 23 (1), 2016 - Conference Reports: Escola Gáucha de Bioinformática, 2015, Porto Alegre / RS.
6. **Pedebos, C.;** Arantes, P. R.; Giesel, G. M.; Verli, H. Structural Glycobiology of the Oligosaccharyltransferase PglB from *Campylobacter lari*. Programa de Resumos da XLIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014, Foz do Iguaçu / PR.
7. Dallagnol, J. C. C.; Arantes, P. R.; **Pedebos, C.;** Braun, R. L.; Duarte, M. E.; Nosedá, M. D.; Ducatti, D. R. B.; Verli, H.; Gonçalves, A. G. Influence of ring conformation on interactions of a carbohydrate based compound in a M1 muscarinic acetylcholine receptor model. 7th Brazilian Symposium on Medicinal Chemistry, 2014, Campos do Jordão / SP.
8. **Pedebos, C.;** Giesel, G. M.; Verli, H. Structural Biology of the Oligosaccharyltransferase PglB from *Campylobacter lari*. Programa de Resumos da XLII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2013, Foz do Iguaçu / PR.
9. **Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H. Atomic Model for Micelles Composed by the Saponin QS-21. Programa de Resumos da XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012, Foz do Iguaçu / PR.
10. 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.
11. **Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H. Structure and Conformation of QS-21 Saponin: Implications for Micelle Formation.

Programa de Resumos da XL Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2011, Foz do Iguaçu / PR.

12. **Pedebos, C.;** Pol-Fachin, L.; Teixeira, C. V.; Verli, H. Caracterização Conformacional da Saponina QS-21 e as Implicações na Formação Micelar. Livro de resumos - XXIII Salão de Iniciação Científica e XX Feira de Iniciação Científica, 2011, Porto Alegre / RS.
13. **Pedebos, C.;** Pol-Fachin, L.; Verli, H. Depiction of Saponins Conformational Ensemble. Programa de Resumos da XXXIX Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2010, Foz do Iguaçu / PR.
14. **Pedebos, C.;** Pol-Fachin, L.; Verli, H. Caracterização Conformacional de Saponinas. Livro de resumos - XXII Salão de Iniciação Científica e XIX Feira de Iniciação Científica, 2010, Porto Alegre / RS.
15. **Pedebos, C.;** Pol-Fachin, L.; Verli, H. Conformational characterization of saponins conformation in pyridine. Programa de Resumos da XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2009, Águas de Lindóia / SP.
16. **Pedebos, C.;** Pol-Fachin, L.; Verli, H. Caracterização Conformacional de Saponinas em Piridina. Livro de resumos - XXI Salão de Iniciação Científica, 2009, Porto Alegre / RS.

WORK PRESENTATIONS

1. **Pedebos, C.;** Pol-Fachin, L.; Verli, H. Caracterização Conformacional de Saponinas em Piridina. Workshop “Fitomedicamentos da química à terapêutica”, 2009, Porto Alegre / RS.

AWARDS

1. 7th Brazilian Symposium on Medicinal Chemistry, 2014: Best Graduate Student Poster Of The Drug Discovery & Development Session, with the poster entitled “Influence of ring conformation on interactions of a carbohydrate based compound in a M1 muscarinic acetylcholine receptor model”.
2. XXII Salão de Iniciação Científica, 2010: Best Work in Session Phytochemistry I - XXII Salão de Iniciação Científica - Health Sciences Area - UFRGS.

CURRENT/ONGOING COLLABORATIONS

1. da Costa, C. T.; **Pedebos, C.**; Verli, H.; Fett-Neto, A. The role of Zn²⁺, dimerization and N-glycosylation in the interaction of Auxin Binding Protein 1 (ABP1) with different auxins. Submitted to *Glycobiology*.
2. Feltes, B. C.; **Pedebos, C.**; Verli, H.; Bonatto, D. Conformational study of the DDB2-DDB1 mutated protein complex in Xeroderma Pigmentosum disease. In preparation for *Nucleic Acids Research*.
3. Chiodi, C.; Ligabue-Braun, R.; **Pedebos, C.**; Verli, H. Structural glycobiology of gangliosides: A Comparative Molecular Mechanics Study of GM1, GD1b and GT1b. In preparation for *Glycoconjugates*.

PARTICIPATION IN EXAMINATION BOARDS

1. **Pedebos, C.**; Stanisçuaski, F. Participation In Board of Juliano de Oliveira Silveira. 2014. Ativação de ureases: inferências evolutivas via filogenia. Course Conclusion Project (Bachelor Degree in Biomedicine) - Universidade Federal do Rio Grande do Sul.

JOURNAL REVIEWER

Revista de Informática Teórica e Aplicada (RITA) - Reviewer

2014 - NOW

PARTICIPATION IN EVENTS

1. XLV Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2016, Natal / RN. (Congress)
2. XLIV Annual Meeting of the Brazilian Society for Biochemistry and Molecular Biology – SBBq and XXIII Congress of the International Union of Biochemistry and Molecular Biology (IUBMB). 2015, Foz do Iguaçu / PR. (Congress)
3. 1ª Escola Gaúcha de Bioinformática, 2015, Porto Alegre / RS.
4. XLIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2014, Foz do Iguaçu / PR. (Congress)
5. XLII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2013, Foz do Iguaçu / PR. (Congress)
6. XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2012, Foz do Iguaçu / PR. (Congress)

-
7. XL Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2011, Foz do Iguaçu / PR. (Congress).
 8. XXIII Salão de Iniciação Científica, UFRGS, 2011, Porto Alegre / RS.
 9. XXXIX Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2010, Foz do Iguaçu / PR. (Congress).
 10. XXII Salão de Iniciação Científica, UFRGS, 2010, Porto Alegre / RS.
 11. XXXVIII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular. 2009, Águas de Lindóia / SP. (Congress).
 12. XXI Salão de Iniciação Científica, UFRGS, 2009, Porto Alegre / RS .
 13. II Semana Acadêmica de Biomedicina da Universidade Federal do Rio Grande do Sul. 2008, Porto Alegre / RS.