

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

INTRODUÇÃO DE INFORMAÇÃO SACARÍDICA EM
ESTRUTURAS CRISTALINAS E SUA IMPLICAÇÃO NO
ENTENDIMENTO DE PROCESSOS PATOLÓGICOS

Liana Guimarães Sachett Petersen

Porto Alegre – Brasil

Setembro de 2014

Introdução de Informação Sacarídica em Estruturas Cristalinas e sua Implicação no Entendimento de Processos Patológicos

Liana Guimarães Sachett Petersen

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

Hugo Verli

Porto Alegre – Brasil

Setembro de 2014

Introdução de Informação Sacarídica em Estruturas Cristalinas e sua Implicação no Entendimento de Processos Patológicos

Liana Guimarães Sachett Petersen

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:

Hugo Verli
Centro de Biotecnologia - UFRGS
Orientador

João Alexandre Ribeiro Gonçalves Barbosa
Instituto de Ciências Biológicas
Universidade de Brasília

Márcia Rosângela Wink
Departamento de Ciências Básicas da Saúde - Bioquímica
Universidade Federal de Ciências da Saúde de Porto Alegre

Gustavo Fioravanti Vieira
Programa de Pós-Graduação em Genética e Biologia Molecular
Universidade Federal do Rio Grande do Sul

Charley Sttats
Centro de Biotecnologia
Universidade Federal do Rio Grande do Sul
[Revisor]

Esta tese foi desenvolvida sob a orientação do Professor Doutor Hugo Verli, com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Pró-Reitoria de Pós-Graduação (UFRGS), como requisito para obtenção do grau de Doutor em Biologia Celular e Molecular, junto ao Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul.

FICHA CATALOGRÁFICA

**INTRODUÇÃO DE INFORMAÇÃO SACARÍDICA EM ESTRUTURAS CRISTALINAS E SUA
IMPLICAÇÃO NO ENTENDIMENTO DE PROCESSOS PATOLÓGICOS**

Rio Grande do Sul, UFRGS, Centro de Biotecnologia, 2014.

Tese: Doutor em Ciências (Biologia Celular e Molecular)

- | | |
|-----------------------|------------|
| 1. Dinâmica Molecular | 2. Fel d 1 |
| 3. FUT8 | 4. C2GnT-L |

I. Hugo Verli

II. Universidade Federal do Rio Grande do Sul – Centro de Biotecnologia

III. Títulos

AGRADECIMENTOS

Ao Prof. Hugo Verli, por ter me recebido junto ao Grupo de Bioinformática Estrutural e ter se dedicado a me orientar.

Ao Programa de Pós-Graduação em Biologia Celular e Molecular, pela oportunidade de realização deste trabalho.

Aos alunos do grupo de Bioinformática Estrutural, pelo companheirismo e convivência, tanto no laboratório quando fora dele.

Ao Luciano e Sílvia, pela competência e atendimento sempre atencioso.

Aos meus pais, pelo apoio e incentivo.

Desde já aos membros da banca, por terem aceitado o convite.

Aos colegas Rodrigo e Cláudia, por tantos conselhos, momentos de descontração e compreensão e presença constante.

Ao meu marido Daniel, pelo apoio incondicional, palavras de incentivo, paciência, amizade, estímulo e exemplo de perseverança.

“Appreciation is a wonderful thing: It makes what is excellent in others belong to us as well.”

(A apreciação é uma coisa maravilhosa: faz o que é excelente nos outros pertencer também a nós).

Voltaire

SUMÁRIO

LISTA DE ABREVIATURAS.....	x
RESUMO	xi
ABSTRACT.....	xii
ÍNDICE DE FIGURAS.....	xiii
1. INTRODUÇÃO.....	15
1.1 CRISTALOGRAFIA DE RAIOS-X.....	15
1.2 ESTRUTURA E SÍNTESE DE GLICANAS	19
1.3 O PAPEL DE GLICANAS NA FUNÇÃO PROTEICA E EM PROCESSOS PATOLÓGICOS	22
1.4 GLICANAS E ALERGENICIDADE.....	25
1.5 FEL D 1, O PRINCIPAL ALÉRGENO DO GATO	27
1.6 GLICANAS E CÂNCER.....	30
1.7 GLICOSILTRANSFERASES	31
1.7.1 FUCOSILTRANSFERASE 8	35
1.7.2 CORE-2 N-ACETIL GLICOSAMINILTRANSFERASE LEUCOCITÁRIA.....	37
1.8 DINÂMICA MOLECULAR.....	40
2. OBJETIVOS.....	44
3. MÉTODOS.....	46
3.1 SISTEMAS ESTUDADOS.....	46
3.2 PROGRAMAS UTILIZADOS	47
3.3 PROTOCOLOS.....	47
3.3.1 DM.....	47
3.3.2 Cálculo semiempírico de Fel d 1 com íons Ca ²⁺	50
3.3.3 Atracamento de FUT8 à glicana-alvo	50
3.3.4 Atracamento de FUT8 à integrina α5β1	51
4. RESULTADOS.....	52
4.1 Artigo de divulgação científica: O Papel de Carboidratos na Função Proteica: o caso das PGHSs	53
4.2 Artigo: The calcium goes meow: effects of ions and glycosylation on Fel d 1, the major cat allergen	65
4.3 Manuscrito: Conformational characterization of fucosyltransferase 8 (FUT8) in solution	95

4.4 Caracterização da glicobiologia estrutura de C2GnT-L.....	114
4.2.1 DM de C2GnT-L livre e complexada aos seus substratos doador e acceptor	114
5. DISCUSSÃO GERAL.....	121
5.1 Preâmbulo.....	121
5.2 Glicobiologia estrutural da Fel d 1.....	121
5.3 Glicobiologia estrutural de FUT8.....	123
5.4 Glicobiologia estrutura de C2GnT-L.....	124
5.5 Impacto de múltiplas simulações por DM.....	125
6. CONCLUSÕES.....	127
7. PERSPECTIVAS.....	129
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	130
9. APÊNDICES.....	142
Apêndice A.....	142
Apêndice B.....	150
10. CURRICULUM VITAE.....	164

LISTA DE ABREVIATURAS

C2GnT-L	Core-2 N-acetil glicosaminiltransferase leucocitária
DM	Dinâmica Molecular
Fuc	Fucose
FUT8	Fucosiltransferase 8
Gal	Galactose
GlcNAc	N-acetil glicosamina
GT(s)	Glicosiltransferase(s)
NMA	Análise de Modos Normais (<i>Normal modes analysis</i>)
NOE	<i>Nuclear Overhauser Effect</i>
PDB	<i>Protein Data Bank</i>
RMN	Ressonância Magnética Nuclear
RMSD	Desvio médio quadrático (<i>Root mean square deviation</i>)
RMSF	Flutuação média quadrática (<i>Root mean square fluctuation</i>)

RESUMO

Sabe-se há algum tempo que a expressão de glicanas muda de acordo com a condição celular. A alteração do processo de glicosilação de proteínas é um processo comum em vários estados fisiopatológicos, sendo observado como um evento frequente em processos alérgicos e em células tumorais. O principal alérgeno do gato, a proteína Fel d 1, é um dímero composto de duas subunidades, sendo cada subunidade composta por uma cadeia α e β . A proteína possui um sítio de N-glicosilação em cada uma de suas subunidades, sendo preenchidos por diferentes glicofomas. Além disso, estudos cristalográficos detectaram três possíveis sítios de ligação a Ca^{2+} em sua estrutura. Diversos tipos de tumores apresentam expressão aumentada de GTs, dentre estas a FUT8 e a C2GnT-L. A FUT8 é uma α -1,6-fucosiltransferase, que transfere uma Fuc do substrato doador (GDP- β -L-fucose) à primeira GlcNAc do núcleo pentassacarídico de uma N-glicana. Esta enzima apresenta expressão aumentada em câncer de pulmão, gástrico e de cólon, além de estar envolvida em outros processos patológicos. A C2GnT-L transfere um resíduo de GlcNAc, em ligação β -1,6, ao núcleo 2 (Core 2) de glicanas O-ligadas, sendo superexpressa em câncer de pulmão, próstata e colorretal e relacionada com progressão tumoral e metástase. Considerando o papel da glicosilação e dos sítios ligadores de Ca^{2+} nos processos alérgicos, assim como a presença das GTs mencionadas em alta quantidade em processos tumorais, bem como a inexistência de dados estruturais e conformacionais destas proteínas e das estruturas sacarídicas em solução, o trabalho tem como objetivo contribuir na caracterização da glicobiologia estrutural destas proteínas. Para tal, foram estudados diversos sistemas da proteína Fel d 1 (PDBID 2ejn) e das enzimas FUT8 (PDBID 2de0) e C2GnT-L (PDBIDs 2gak e 2gam) por DM em solução aquosa em diferentes condições. Após os experimentos, pode-se observar que, para Fel d 1, apenas o íon Ca^{2+} central está ligado à proteína, enquanto os laterais ficam livres. A Fel d 1 com glicosilação reduzida apresentou um aumento de flexibilidade em duas hélices em uma das subunidades da Fel d 1, devido a desenovelamento das mesmas. Já a proteína completamente glicosilada apresentou uma redução no tamanho da cavidade de uma das subunidades. No estudo da FUT8, observou-se que o domínio N-terminal é o principal componente da flexibilidade proteica desta proteína. Ainda, foi possível caracterizar, através de atracamento molecular, a região de interação da FUT8 com sua glicana-alvo e com uma de suas proteínas-alvo, a integrina $\alpha 5\beta 1$. Os resultados para C2GnT-L demonstraram que o sítio caracterizado para o substrato aceptor pode na verdade estar ocupado pelo substrato doador desta enzima, já que este permanece no local durante a DM, ao contrário do substrato aceptor. Espera-se que este trabalho contribua na compreensão do funcionamento destas proteínas, possibilitando futuros estudos acerca de seus papéis em enfermidades e, conseqüentemente, para novas estratégias terapêuticas.

ABSTRACT

We know for a while that glycan expression varies according to the cellular condition. Alterations in protein glycosylation is a common process in several physical and pathological processes, and are frequently observed in allergy and tumorous cells. The major cat allergen, Fel d 1, is a dimer composed by two subunits, each subunit containing one α and β chain. This protein contains one N-glycosylation site in each subunit, which are filled by different glycoforms. In addition, crystallographic studies detected three possible Ca^{2+} binding sites in its structure. Several types of tumors have higher expression of GTs, like FUT8 and C2GnT-L. FUT8 is an α -1,6-fucosyltransferase, transferring one Fuc from GDP-b-L-fucose to the inner most GlcNAc in the pentasaccharidic core of an N-glycan. This enzyme is highly expressed in lung, gastric and colon cancer and is involved in other pathological processes. C2GnT-L transfers a GlcNAc residue, in a β -1,6 linkage, to the Core 2 of O-linked glycans and is overexpressed in lung, prostate and colon-rectal cancer, related to tumoral progression and metastasis. Considering the role of glycosylation and Ca^{2+} binding sites in allergy, and considering the presence of the GTs mentioned above in high amounts in tumoral processes, as well as the inexistence of structural data concerning these proteins in solution, this study intends to contribute in characterizing the structural glycobiology of these proteins. In this matter, we studied several systems containing Fel d 1 (PDBID 2ejn), FUT8 (PDBID 2de0) and C2GnT-L (PDBID 2gak and 2gam) by MD in aqueous solution, under different conditions. We were able to observe, after the experiments, that in Fel d 1, only the central Ca^{2+} ion is bound to the protein, while the lateral ions are free. Fel d 1 containing reduced glycosylation form presented larger flexibility in two helices, in one of the subunits, due to their unfolding. Fully glycosylated Fel d 1 presented reduced cavity size in one of the subunits. In FUT8 study, we observed that the N-terminal domain is the principal component of protein flexibility. We were also able to characterize, using molecular docking, the sites of interaction of FUT8 with its target glycan and one of its target protein, integrin $\alpha 5\beta 1$. Results for C2GnT-L showed that the characterized acceptor substrate site might, in fact, be occupied by the donor substrate, since the donor substrate remains in the site during MD, and the acceptor substrate does not. We expect that the refinement of these crystallographic structures, by introducing the saccharidic information, will contribute in comprehending the proteins functioning, and will further possibilitate studies into its roles in infirmities and, in this matter, to new therapeutic strategies.

ÍNDICE DE FIGURAS

Figura 1. Passos na resolução de uma estrutura cristalográfica (PETSKO E RINGE, 2009).	16
Figura 2. Mapas de densidade eletrônica com um peptídeo (correspondente a Ser-Val-Val-Met-Thr-Ile) sobreposto. Em A), a resolução é de 3Å e pode-se determinar a posição aproximada das cadeias laterais. Em B), com 2 Å de resolução, as cadeias laterais estão bem delineadas. Em C) temos 1Å de resolução e os átomos são identificáveis (PETSKO E RINGE, 2009).....	17
Figura 3. Diversidade estrutural das glicanas. Com base na estrutura do esqueleto sacarídico, as glicanas podem ser amplamente divididas entre lineares (glicosaminoglicanos) e ramificadas (glicanas N- e O-ligadas). (A) Glicanas O-ligadas, unidas a proteínas por ligações covalentes (tipicamente serina ou treonina) subagrupadas de acordo com o núcleo (“core”) estrutural. (B) Glicosilações N-ligadas, onde a glicana está tipicamente unida covalentemente a uma asparagina em uma proteína contendo a sequência consenso Asn-X-Ser/Thr, onde X pode ser qualquer aminoácido exceto prolina. Glicanas N-ligadas geralmente apresentam estruturas maiores e mais complexas que as O-ligadas, mas possuem menor variedade. Todas as N-glicanas possuem um núcleo pentassacarídico comum (GlcNAc ₂ Man ₃) e podem ser classificadas em três tipos: oligomanose (high-mannose), complexo ou híbrido. (C) GAGs, o principal grupo de glicanas lineares, correspondente a unidades dissacarídicas sulfatadas repetidas de um ácido urônico ligada a uma hexosamina, que quando ligados a proteínas chamam-se proteoglicanos. Três exemplos de GAGs são apresentados: heparan sulfato (HS), condroitin sulfato (CS), e dermatan sulfato (ROBINSON ET AL., 2012).	20
Figura 4. Comparação entre um modelo e múltiplos modelos de glicoproteínas, com destaque para a variação da porção carboidrato (Adaptado de POL-FACHIN E VERLI, 2011; por Pablo Arantes).	23
Figura 5. Sequência de eventos na reação de hipersensibilidade imediata. A de hipersensibilidade imediata se inicia pela introdução de uma alérgeno, que estimula a reação de células T _H 2 e produção de IgE, que sensibiliza mastócitos por ligar ao receptor Fc, que por sua vez secreta mediadores responsáveis pelas reações patológicas (ABBAS, 2012).	25
Figura 6. Representação esquemática da distribuição de Fel d 1 no gato doméstico, Felis domesticus.	28
Figura 7. Estrutura do tetrâmero de Fel d 1 com os sítios de ligação a Ca ²⁺ , conforme o PDB ID 2ejn. (Subunidades heterodiméricas representadas pelas cores amarela e azul e íons Ca ²⁺ representados por esferas laranjas).	29
Figura 8. A) Mecanismo de deslocamento duplo, envolvendo um intermediário covalente glicosil-enzima. B) Inversão por SN2, com ataque em C1 pelo lado oposto pela hidroxila aceptor. Mecanismos alternativos para retenção: C) Mecanismo ortogonal, com ataque nucleofílico em C1 pelo aceptor; D) Mecanismo SNi; E) Mecanismo SN1 (SCHUMAN ET AL., 2013).....	32
Figura 9. A) Enovelamento GT-A (PDB ID 1h7l) e B) Enovelamento GT-B (PDB ID 1nlm).....	33

Figura 10. A) N-ligadas e B) O-ligadas. (Adaptado de FUSTER & ESKO, 2005).....	34
Figura 11. Reação de α 1,6 fucosilação catalisada pela FUT8 (transferência de uma fucose, em ligação α 1,6, para a primeira GlcNAc de uma glicana N-ligada).....	35
Figura 12. A) Estrutura cristalográfica de FUT8 humana (PDB ID 2de0). Domínio N-terminal (coiled-coil) em cinza. Domínio catalítico colorido de acordo com a estrutura secundária (α -hélices em vermelho, folhas- β em amarelo e alças em verde). E domínio C-terminal SH3 em azul. B) Domínio catalítico, destacando a estrutura de sanduíche $\alpha/\beta/\alpha$ (α -hélices em azul ciano, folhas- β em roxo/rosa e alças em salmão). C) Domínio catalítico, destacando a estrutura tipo Rossmann (α -hélices em azul ciano, folhas- β em roxo/rosa e alças em salmão).	36
Figura 13. Reação catalisada por C2GnT-L (transferência de GlcNAc, em ligação 1,6, para uma estrutura Core 1, formando a estrutura Core 2).	37
Figura 14. Estrutura cristalográfica de C2GnT-L (PDB ID 2gak). Resíduos 38 a 121 (α -hélices) coloridos em cinza. O restante corresponde ao domínio catalítico (resíduos 122 a 428), coloridos de acordo com a estrutura secundária (α -hélices em vermelho, folhas- β em amarelo e alças em verde).....	39
Figura 15. Funções de energia que compõem campos de força empregados em dinâmica molecular (LIGABUE-BRAUN, TESE DE DOUTORADO, 2014).....	42
Figura 16. Estrutura da glicana alvo da fucosilação, posteriormente submetida ao atracamento molecular.	51
Figura 17. RMSD para C2GnT-L livre utilizando a estrutura de PDB 2gak (A), para C2GnT-L livre utilizando a estrutura de PDB 2gam (B), para C2gnT-L complexada a seu substrato acceptor, Gal- β -(1-3)-GalNAc, obtida do PDB 2gam (C), e para C2GnT-L complexada ao seu substrato doador, UDP-GlcNAc (D). As cores preta e vermelha correspondem respectivamente ao domínios N-terminal e catalítico.	114
Figura 18. RMSF geral (preto) e análise de componentes principais (vetor 1 vermelho e vetor 2 azul) para C2GnT-L livre utilizando a estrutura de PDB 2gak (A), para C2GnT-L livre utilizando a estrutura de PDB 2gam (B), para C2gnT-L complexada a seu substrato acceptor, Gal- β -(1-3)-GalNAc, obtida do PDB 2gam (C), e para C2GnT-L complexada ao seu substrato doador, UDP-GlcNAc (D).	115
Figura 19. Variação da estrutura secundária durante o tempo de simulação por resíduo (conforme legenda de cores).....	116
Figura 20. Visualização das estruturas iniciais (0ns, vermelho) e finais (100ns, azuis) de cada sistema. A região N-terminal está demonstrada em cores claras (salmão para 0ns e azul claro para 100ns). A região de maior flexibilidade (resíduos 215 a 220) está destacada pela superfície azul nas estruturas finais.	117
Figura 21. Visualização da estrutura inicial (0ns, vermelha) e final (100ns, azul) dos sistemas com substrato. Posição inicial (A) e final (B) do substrato acceptor (esferas) e posição inicial (C) e final (D) do substrato doador (esferas).	118
Figura 22. A) Conformação cristalográfica do substrato acceptor (Gal- β -(1-3)-GalNAc). B) Mapa de energia para este dissacarídeo após metadinâmica (regiões azuis = menor energia; regiões vermelhas = maior energia). C) Gráfico com os picos de prevalências conformacionais para o dissacarídeo após a DM em solução.....	119
Figura 23. RMSD para os domínios N-terminal (coiled-coil, preto), catalítico (vermelho) e SH3 (azul) dos três sistemas de FUT8 complexada ao substrato GDP- β -L-fucose.	126

1. INTRODUÇÃO

1.1 CRISTALOGRAFIA DE RAIOS-X

A cristalografia de macromoléculas iniciou há mais de meio século, com a estrutura proteica da mioglobina, com 6Å de resolução (KENDREW ET AL, 1938). A partir de então, o *Protein Data Bank* (PDB) (BERNSTEIN ET AL., 1977; BERMAN ET AL., 2000) foi estabelecido como o único repositório fonte de estruturas cristalográficas (e, mais tarde, de outros modelos estruturais obtidos por RMN, microscopia eletrônica, etc.) da comunidade científica (WLODAWER ET AL., 2008).

A determinação de estruturas atômicas de macromoléculas pode ser feita por diversas técnicas, mas a maioria das estruturas é obtida pela cristalografia de raios-X ou RMN. Ambas as técnicas informam a posição relativa dos átomos da molécula (coordenadas atômicas). Na cristalografia, a nuvem eletrônica dos átomos é a responsável pela dispersão dos raios-X. Já na RMN são medidas as interações dos núcleos atômicos. Desta maneira, a cristalografia de raios-X fornece a posição de todos os átomos, exceto hidrogênios, enquanto a RMN fornece a posição de todos os átomos incluindo os hidrogênios. A escolha do método depende de uma variedade de fatores, incluindo peso molecular, solubilidade e facilidade de cristalização. Para proteínas ou complexos proteicos com peso molecular acima de 50-100 kDa, deve-se usar cristalografias de raios-X. Para aquelas proteínas de difícil cristalização e capazes de se dissolverem em altas concentrações sem agregar, deve-se usar RMN. Em muitos casos, podem ser usadas ambas as técnicas de forma complementar, pois a cristalografia de raios-X fornece imagens estáticas e a RMN pode ser usada para estudar a flexibilidade de proteínas na solução (PETSKO E RINGE, 2009).

A obtenção de proteínas para cristalografia pode ser feita de diversas maneiras. Algumas proteínas podem ser obtidas diretamente de sua fonte natural, mas para que isso aconteça ela deve estar em alta concentração, o que não acontece na maioria dos casos, pois uma proteína costuma estar distribuída em diversos tecidos. Além disso, se existirem isoformas de modificações pós-traducionais, essa heterogeneidade da amostra causará um impacto na pureza

da mesma e isso influenciará posteriormente a formação de cristais. A produção de proteínas puras geralmente é feita através de métodos de expressão heteróloga e a escolha do organismo de expressão deve levar em conta a necessidade de modificações pós-traducionais (por exemplo, um sistema bacteriano não é capaz de glicosilar proteínas de eucariotos). Desta maneira, pode-se ter limitações estabelecidas pelo sistema de expressão. Mas além de necessitarmos de uma proteína pura, o passo limitante na resolução de uma estrutura cristalográfica proteica é a proteína ser cristalizada. Esse passo é essencial e existem proteínas difíceis de cristalizar (especialmente proteínas de membrana). Muitas vezes torna-se necessário utilizar alta concentração de agentes precipitantes, alterar o pH, a temperatura, a pressão, entre outras mudanças (MALUF ET AL., 2014). Esse ambiente cristalino muitas vezes acaba se tornando muito diferente do ambiente biológico da proteína, o que pode resultar em erros se não forem feitas as corretas interpretações.

Após a obtenção do cristal, a difração de raios-X é computada. A fonte de raios-X geralmente é um sincrotron. Os cristais são bombardeados com raios-X que se espalham e são capturados como um padrão de difração. O que se mede experimentalmente é a amplitude e posição das ondas de raios-X difratadas. A estrutura pode ser reconstruída pelo somatório destas ondas, utilizando referências (ou fases), a partir de átomos marcados no cristal (por exemplo, marcação com átomos pesados). Isso leva a obtenção de mapas de densidade eletrônica. Uma estrutura química é então encaixada neste mapa e o modelo é ajustado e refinado num processo iterativo até que o padrão de difração calculado e o medido experimentalmente equivalham (Figura 1).

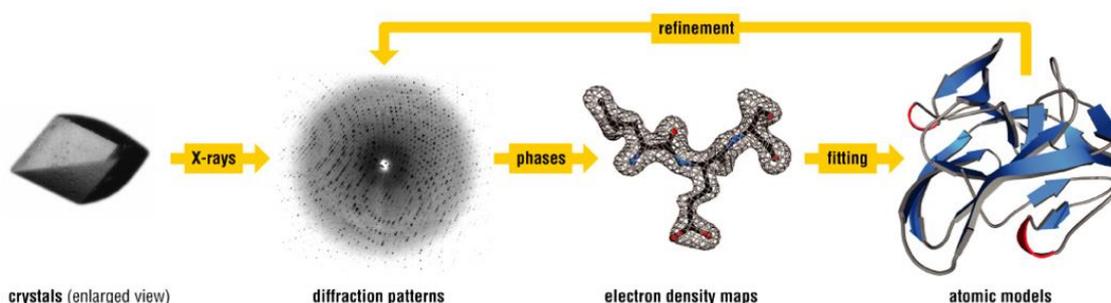


Figura 1. Passos na resolução de uma estrutura cristalográfica (PESKO E RINGE, 2009).

A interpretação do mapa de densidade eletrônica sofre interferência da resolução, que indica o nível de detalhamento com o qual a proteína foi determinada. Diferentes níveis de resolução determinam tipos de informação distintas. Numa resolução de 5Å, identifica-se a topologia da molécula e elementos de estrutura secundária. Com 3Å as cadeias laterais de alguns aminoácidos são interpretáveis. Com 2,5Å as cadeias laterais de todos os aminoácidos são interpretáveis. Para reconhecer átomos individuais, necessita-se uma resolução de 1,5Å. E com 1Å de resolução, os tipos de átomos são identificáveis. Um exemplo de mapa de densidade eletrônica com diferentes resoluções é apresentado na [Figura 2](#).



[Figura 2](#). Mapas de densidade eletrônica com um peptídeo (correspondente a Ser-Val-Val-Met-Thr-Ile) sobreposto. Em A), a resolução é de 3Å e pode-se determinar a posição aproximada das cadeias laterais. Em B), com 2 Å de resolução, as cadeias laterais estão bem delineadas. Em C) temos 1Å de resolução e os átomos são identificáveis (PETSKO E RINGE, 2009).

Diversos fatores influenciam a interpretação de um mapa de densidade eletrônica, já que ele corresponde a uma média das posições atômicas ao longo de todas as células unitárias que formam o cristal. Se a densidade eletrônica não é bem definida ou se houverem moléculas em diferentes orientações, a interpretação do mapa torna-se desafiadora. A cadeia lateral de um resíduo de aminoácido num peptídeo pode adotar mais de uma conformação e se existirem muitas conformações distintas a densidade eletrônica para esta cadeia lateral não será distinguível e aparecerá como um ruído no mapa (WLODAWER ET AL., 2008; MALUF ET AL., 2014). É muito comum acontecer esse tipo de ruído no caso de regiões desordenadas ou regiões altamente flexíveis, como alças. Isso também é comum para estruturas sacarídicas em glicoproteínas, que geralmente estão incompletas no PDB, pois a alta flexibilidade das glicanas impede tanto a cristalização quanto a aquisição de densidade eletrônica (WOODS, 1998). Mesmo

quando conseguem ser obtidas por cristalografia de raios-X, estas estruturas sacarídicas são de baixa qualidade, e estima-se que um terço dos dados para carboidratos encontrados no PDB contém erros significativos de estereoquímica, nomenclatura e concordância com mapas de densidade eletrônica (LÜTEKE ET AL., 2004; CRISPIN ET AL., 2007; BERMAN ET AL., 2007). Mapas de densidade eletrônica podem ser fracos também no caso de ligantes, quando este interage com apenas algumas moléculas do cristal. Essa ocupação parcial torna difícil a interpretação e modelagem destes ligantes (MALUF ET AL., 2014).

Os modelos estruturais construídos com base na cristalografia de raios-X devem ser precisos, e, para isso, alguns métodos de refinamento devem ser aplicados. Uma das estratégias empregadas é o alinhamento sistemático entre o modelo estrutural e a densidade eletrônica, feito por ciclos contínuos até que ocorra convergência de dados. Para auxiliar o refinamento, são aplicadas restrições estereoquímicas para orientar o grau de liberdade conformacional dos átomos nas tentativas de modelá-los. Estas restrições são baseadas em dados de altas resoluções de pequenas moléculas e garantem a não violação de estruturas permitidas para diferentes grupos químicos, bem como impede que eles adotem conformações de alta energia. Outra estratégia envolve o conhecimento das características geométricas dos aminoácidos e das estruturas secundárias (distâncias, ângulos de ligação e ângulos de diedro), que servem de guias para a avaliação dos modelos (MALUF ET AL., 2014).

Cabe ressaltar que, para carboidratos, a descrição estrutural por RMN também encontra dificuldades, pois a alta flexibilidade das glicanas produz um conjunto complexo de conformações, com a possibilidade de coexistirem múltiplos conformeros simultaneamente em solução. Além disso, geralmente tem-se um baixo número ou ausência de sinais de NOE. Uma determinação conformacional precisa de carboidratos, utilizando essa metodologia, só é possível se forem obtidos três ou mais sinais de NOE inter-resíduos em uma ligação glicosídica, o que, na maioria das vezes, não permite a determinação conformacional de oligossacarídeos maiores que uma dezena de resíduos (CUMMING E CARVER, 1987; WOODS, 1998; WORMALD ET AL., 2002). A espectrometria de massas, por sua vez, é a técnica que melhor descreve as estruturas dos oligossacarídeos, embora não ofereça informações sobre a

organização espacial, tridimensional, desta classe de biomoléculas (PÉREZ E MULLOY, 2005).

1.2 ESTRUTURA E SÍNTESE DE GLICANAS

Os termos glicana e polissacarídeo são definidos pela IUPAC como sinônimos que significam “compostos que consistem de um grande número de monossacarídeos unidos por ligações glicosídicas” (IUPAC, 1997). Na prática, o termo glicana pode também ser usado para definir a porção carboidrato de um glicoconjugado, como glicoproteínas, glicolípídeos e proteoglicanos (DWEK, 1996). Ao contrário de proteínas, as glicanas não estão codificadas diretamente no genoma e sua síntese não segue molde ou referência direta, mas sim um processo complexo específico. Embora o genoma não codifique diretamente uma sequência sacarídica, ele determina a expressão de diversas enzimas que sintetizam carboidratos, realizam a sua ligação uns aos outros, modificam monossacarídeos adicionando ou removendo substituintes (VERLI, 2014). A principal família de enzimas envolvidas nestes processos são as glicosiltransferases, como veremos mais tarde no item 1.7, além de glicosidases, ambas com expressão e atividade dependentes do tecido ou do estado metabólico (ROBINSON ET AL., 2012).

A complexidade de estudar glicanas pode ser descrita ao compararmos estas moléculas com proteínas. Os carboidratos apresentam uma grande diversidade em comparação aos 20 aminoácidos codificados no genoma. Mais de 100 possíveis unidades monossacarídicas já foram observadas em biomoléculas. Em uma analogia à ligação peptídica, carboidratos ligam-se entre si ou a outras moléculas por ligações glicosídicas. Mas, enquanto aminoácidos possuem somente um grupo amino e um grupo ácidos carboxílico em seu esqueleto peptídico, de forma que só um tipo de ligação peptídica é possível entre dois resíduos (como também ocorre com nucleotídeos), na ligação glicosídica entre dois monossacarídeos existem diversas possibilidades. A ligação glicosídica é formada pela reação entre dois grupos hidroximetileno (CHOH), e cada monossacarídeo possui vários destes grupos. Desta maneira, múltiplas ligações

entre dois monossacarídeos consecutivos são possíveis, criando assim um espectro complexo de possibilidades (ROBINSON ET AL., 2012; VERLI, 2014).

As glicanas geralmente são formadas por ligações O-glicosídicas entre monossacarídeos. Elas podem ser homo ou heteropolímeros, lineares ou ramificadas. Glicanas ramificadas são comuns em glicosilações N- e O-ligadas em glicoproteínas e glicolípídeos, enquanto glicanas lineares são comuns em glicosaminoglicanos (polímeros geralmente sulfatados) (ROBINSON ET AL., 2012). Esta diversidade estrutural pode ser observada na [Figura 3](#).

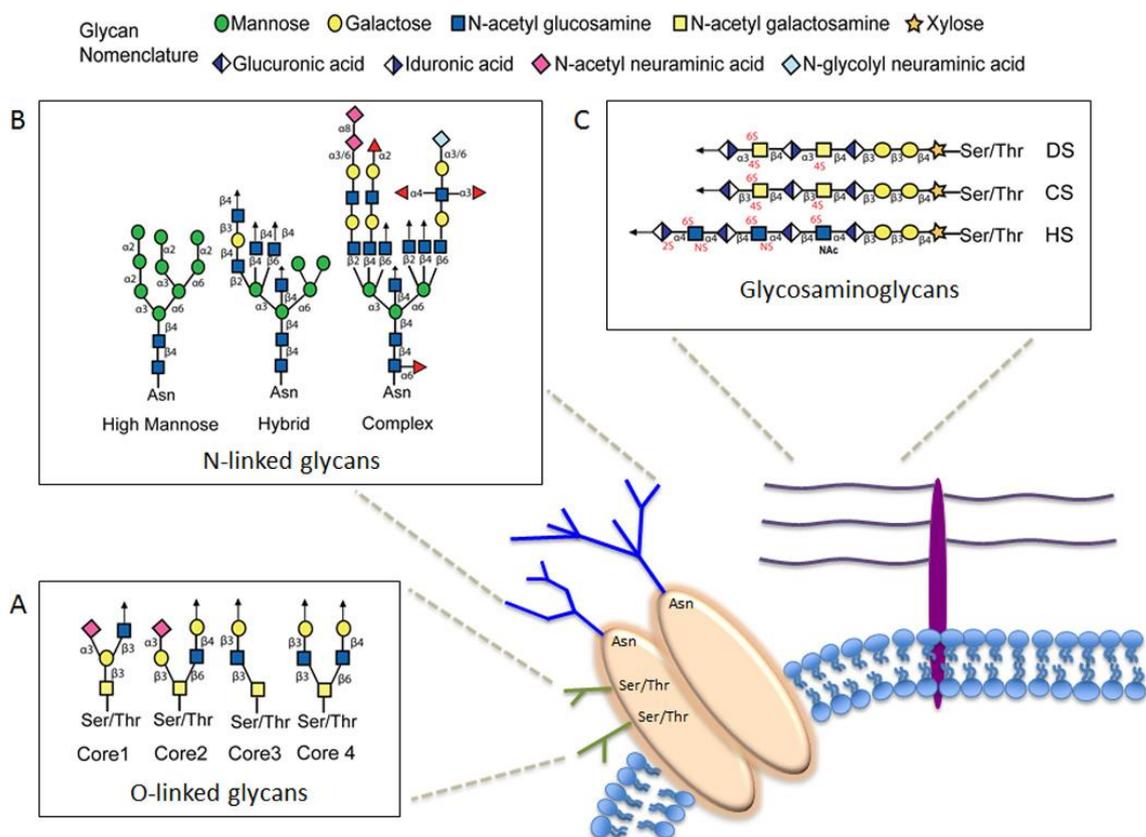


Figura 3. Diversidade estrutural das glicanas. Com base na estrutura do esqueleto sacarídico, as glicanas podem ser amplamente divididas entre lineares (glicosaminoglicanos) e ramificadas (glicanas N- e O-ligadas). (A) Glicanas O-ligadas, unidas a proteínas por ligações covalentes (tipicamente serina ou treonina) subagrupadas de acordo com o núcleo (“core”) estrutural. (B) Glicosilações N-ligadas, onde a glicana está tipicamente unida covalentemente a uma asparagina em uma proteína contendo a sequência consenso Asn-X-Ser/Thr, onde X pode ser qualquer aminoácido exceto prolina. Glicanas N-ligadas geralmente apresentam estruturas maiores e mais complexas que as O-ligadas, mas possuem menor variedade. Todas as N-glicanas possuem um núcleo pentassacarídico comum (GlcNAc₂Man₃) e podem ser classificadas em três tipos: oligomanose (*high-mannose*), complexo ou híbrido. (C) GAGs, o principal grupo de glicanas lineares, correspondente a unidades dissacarídicas sulfatadas repetidas de um ácido urônico ligada a uma hexosamina, que quando ligados a proteínas chamam-se proteoglicanos. Três exemplos de GAGs são apresentados: heparan sulfato (HS), condroitin sulfato (CS), e dermatan sulfato (ROBINSON ET AL., 2012).

Como visto na [Figura 3](#), glicanas O-ligadas ([Figura 3A](#)) são unidas a proteínas tipicamente a resíduos serina ou treonina, e esta classe de glicanas é classificada de acordo com o seu núcleo (“core”) estrutural (ROBINSON ET AL., 2012). A O-glicosilação ocorre no complexo de Golgi e é um evento inteiramente pós-traducional. Na N-glicosilação ([Figura 3B](#)) a glicana liga-se tipicamente a uma asparagina em uma proteína contendo a sequência consenso Asn-X-Ser/Thr, onde X pode ser qualquer aminoácido exceto prolina. Glicanas N-ligadas geralmente apresentam estruturas maiores e mais complexas que as O-ligadas, mas possuem menor variedade. A N-glicosilação inicia no retículo endoplasmático (RE), onde é montado um oligossacarídeo precursor, que será então modificado no complexo de Golgi. A síntese do oligossacarídeo precursor no RE é geralmente realizada co-traducionalmente, uma vez que as proteínas são importadas para o RE antes da síntese completa da cadeia polipeptídica, e algumas proteínas inclusive necessitam da N-glicosilação para seu correto enovelamento (QSABA, 2000). Todas as N-glicanas possuem um núcleo pentassacarídico comum (GlcNAc₂Man₃) e podem ser classificadas em três tipos: oligomanose (high-mannose), complexo ou híbrido (DWEK, 1996; ROBINSON ET AL., 2012). Glicanas tipo oligomanose contém apenas resíduos de manose (Man) ligados ao núcleo pentassacarídico e, portanto, não sofrem a adição de novos carboidratos no aparelho de Golgi. Glicanas tipo complexo não possuem nenhum resíduo de Man além das 3 do núcleo pentassacarídico, mas possuem ramificações com resíduos de N-acetil gliconsamina (GlcNAc) no seu terminal redutor ligadas ao núcleo trimanosil, podendo conter mais do que as duas GlcNAc originais, bem como um número variável de resíduos de galactose (Gal), ácido siálico e, em alguns casos, fucose (Fuc). As glicanas tipo complexo apresentam a maior variação estrutural dentre os subgrupos, resultando principalmente das combinações de diferentes números de ramificações e variações nos monossacarídeos nas cadeias externas. E, por fim, N-glicanas tipo híbrido têm características tanto de glicanas tipo oligomanose quanto de glicanas tipo complexo, com um ou dois resíduos de Man ligados ao núcleo pentassacarídico, como nas glicanas tipo oligomanose, e também com uma ou duas ramificações contendo outros monossacarídeos, como nas glicanas tipo complexo, ligadas a este núcleo (DWEK, 1996).

Desta maneira, temos os diferentes tipos de monossacarídeos, os vários padrões de ramificação, as múltiplas possibilidades de ligação entre monossacarídeos, os diferentes sítios de ligação e ainda a biossíntese das glicanas, que não segue um guia predefinido. Tudo isso demonstra a alta complexidade estrutural das glicanas, que resulta em uma diversidade de estruturas maior do que aquela encontrada em proteínas e ácidos nucleicos.

1.3 O PAPEL DE GLICANAS NA FUNÇÃO PROTEICA E EM PROCESSOS PATOLÓGICOS

As glicoproteínas são um componente majoritário da superfície externa das células mamíferas e representam a maioria das proteínas secretadas em eucariotos (DWEK, 1996). As glicanas ligadas a esses polipeptídios são capazes de modificar diversas propriedades destas moléculas, como estabilidade, enovelamento, solubilidade, conformação, promover resistência a ataques proteolíticos, que assim podem influenciar o papel biológico destas moléculas (VARKI, 1993; WYSS E WAGNER, 1996; TAMS ET AL., 1999; PRICE ET AL., 2003; SHENTAL-BECHOR E LEVY, 2008; HANSON ET AL., 2009; SKROPETA, 2009). Adicionalmente, a fração carboidrato de uma glicoproteína também pode participar diretamente em eventos de reconhecimento, como na adesão entre células, interação com patógenos, sinalização celular, interação com receptores, endocitose (VARKI, 2007; OHTSUBO E MARTH, 2006), bem como no reconhecimento por selectinas e integrinas em situações de infecções, que leva ao recrutamento de linfócitos (ALON E ROSEN, 2007).

O grande tamanho dos carboidratos é provavelmente um dos fatores mais significativos na modificação das propriedades das proteínas as quais eles estão ligados. Grandes áreas de uma proteína podem ser protegidas por oligossacarídeos relativamente pequenos (Figura 4), o que pode permitir que eles cubram áreas funcionalmente importantes e modulem interações de glicoconjugados com outras moléculas (DRICKAMER E TAYLOR, 1993; PAREKH ET AL., 1989; DWEK, 1996).

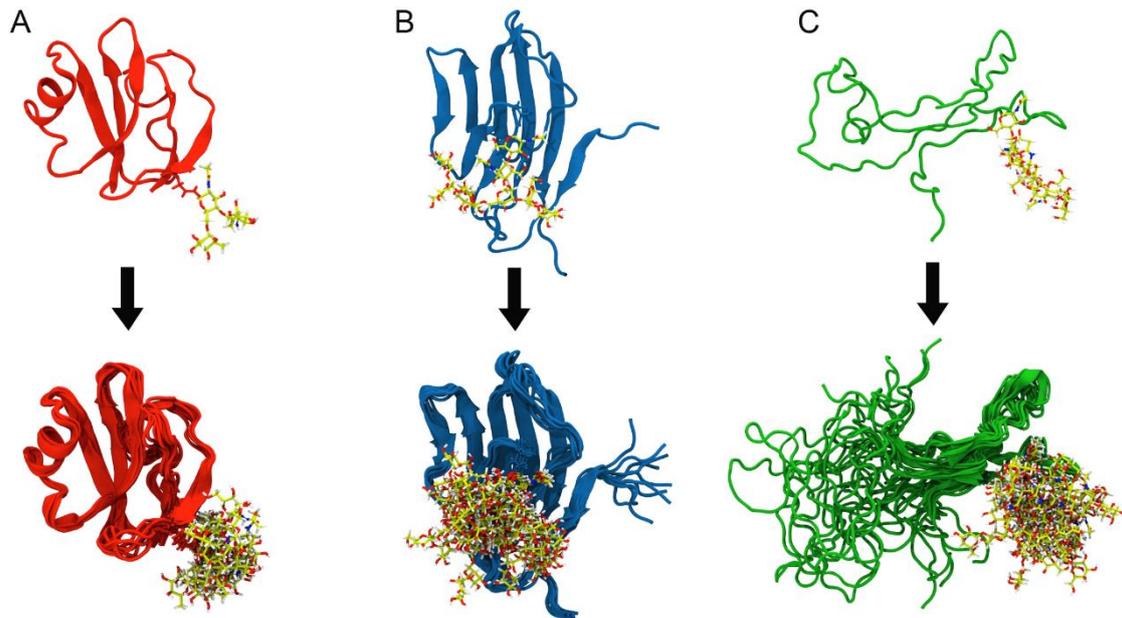


Figura 4. Comparação entre um modelo e múltiplos modelos de glicoproteínas, com destaque para a variação da porção carboidrato (Adaptado de POL-FACHIN E VERLI, 2011; por Pablo Arantes).

Diversos processos patológicos envolvem alterações em glicanas. Por exemplo, pacientes com artrite reumatoide apresentam o mesmo tipo de glicana bi-ramificada em Imunoglobulina G (IgG) sérica que indivíduos normais, mas em proporções muito diferentes, havendo uma diminuição de Gal e um aumento correspondente de GlcNAc no término das ramificações (DWEK, 1996).

Mudanças na glicosilação também estão associadas com infecção e inflamação. Um exemplo são infecções intestinais. Todas as células, incluindo as do trato gastrointestinal, são revestidas por glicocálix, que é composto de glicolipídeos, glicoproteínas e proteoglicanos. A mucosa gastrointestinal possui uma camada que é secretada e outra que fica firmemente aderida, e as glicanas secretadas formam uma camada que serve como barreira e sensor físico-químico, composta de mucinas (que são glicoproteínas) e peptídeos associados. Numa infecção gastrointestinal, os patógenos colonizam e invadem o epitélio do hospedeiro usando interações proteína-proteína e glicana-lectina. Durante o processo de colonização e infecção, ocorre a alteração do padrão de glicosilação tanto do patógeno quanto do hospedeiro, em resposta à presença um do outro (MORAN ET AL., 2011).

Desordens de glicosilação também estão associadas à fraqueza muscular na síndrome miastênica/miopática. Isso se deve, entre outros fatores, a

alterações no gene que codifica a enzima ALG2, uma α -1,3-manosiltransferase que catalisa a transferência da segunda manose do núcleo pentassacarídico de N-glicanas (MONIES ET AL., 2014).

Outra desordem de glicosilação com consequências fisiológicas e patológicas é a alteração em canais iônicos. Os canais iônicos voltagem-dependentes são proteínas transmembrana que regulam a excitabilidade elétrica em vários tecidos, como nervos, músculos e coração. Estes canais podem ser glicosilados e a alteração no padrão de glicosilação pode afetar a função dos mesmos, levando a desordens neurológicas e neuromusculares. Estudos recentes com foco no padrão de sialilação de canais sódio e potássio detectaram que a presença de ácidos siálicos nas glicanas O- e N-ligadas altera o mecanismo dos canais e causa mudanças conformacionais nos domínios sensíveis à voltagem, devido às cargas negativas do ácido siálico (BAYCIN-HIZAL ET AL., 2014).

Outras doenças estão associadas a problemas na glicosilação, como a doença de Tay-Sachs, onde um defeito na enzima hexosaminidase A, decorrente de uma mutação no gene que a codifica, provoca um acúmulo de um gangliosídeo (GM2), levando à neurodegeneração. Já a deficiência conjunta nas hexosaminidases A e B leva a doença de Sandhoff, que possui um fenótipo semelhante à doença de Tay-Sachs. Na doença de Gaucher, ocorre deficiência numa enzima glicocerebrosidase (ou β -glicosidase ácida), que realiza a digestão celular de um glicolípido, o glicocerebrosídeo. Essa deficiência causa o acúmulo lisossomal deste glicolípido, levando ao inchaço de órgãos como fígado, baço e à fragilidade óssea (OHTSUBO E MARTH, 2006).

Glicanas e glicoproteínas são também essenciais para o funcionamento do sistema imune. Um dos exemplos mais conhecidos é a família das selectinas, que são glicoproteínas de superfície encontradas em leucócitos que se ligam a glicanas endoteliais próximas a sítios de inflamação e infecção e permitem o extravasamento aos tecidos (JOHNSON ET AL., 2013). O envolvimento de glicanas na resposta imune, na alergia e no câncer será descrito de forma mais detalhada nos próximos tópicos.

1.4 GLICANAS E ALERGENICIDADE

A alergia é uma resposta exagerada do sistema imunológico a uma substância estranha ao organismo, ou seja, uma hipersensibilidade imunomediada a um estímulo externo específico. Todas as reações alérgicas, embora apresentem grandes diferenças nos tipos de antígenos que desencadeiam estas reações, apresentam algumas características em comum. Primeiramente ocorre ativação de células T_H2 e produção de anticorpos IgE. Isso ocorre por uma sequência de eventos, iniciando na exposição ao antígeno, ativação de células T_H2 e células B específicas para o antígeno, produção de anticorpo IgE, ligação do anticorpo aos receptores Fc de mastócitos, ativação dos mastócitos pela reexposição ao antígeno, resultando da liberação de mediadores a partir destes mastócitos e subsequente reação patológica ([Figura 5](#)). Este processo é chamado de sensibilização.

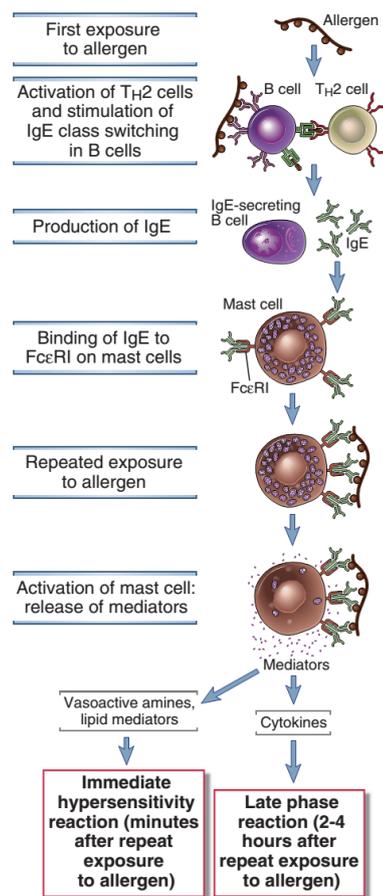


Figura 5. Sequência de eventos na reação de hipersensibilidade imediata. A de hipersensibilidade imediata se inicia pela introdução de um alérgeno, que estimula a reação de células T_H2 e produção de IgE, que sensibiliza mastócitos por ligar ao receptor Fc, que por sua vez secreta mediadores responsáveis pelas reações patológicas (ABBAS, 2012).

Os antígenos que causam a hipersensibilidade imediata (que é a resposta típica da ativação de células T_H2) são chamados alérgenos, e são geralmente proteínas ou moléculas que modulam proteínas. As citocinas produzidas pelas células T_H2 são responsáveis por várias características da hipersensibilidade imediata. As manifestações clínicas e patológicas da hipersensibilidade imediata consistem em reações vasculares e de músculos lisos que se desenvolvem rapidamente após repetidas exposições ao alérgeno e então causam uma reação inflamatória tardia. Todas estas reações são iniciadas por ativação de mastócitos (presentes em tecidos conectivos e sob o epitélio) mediada por IgE, mas diferentes mediadores são responsáveis pelos componentes da fase tardia (que é a resposta clássica da ativação de células T_H1).

A hipersensibilidade imediata se manifesta de diferentes maneiras dependendo do tecido afetado, incluindo constrição brônquica, congestão nasal, dor abdominal, diarreia, erupções cutâneas e até choque sistêmico. Na reação sistêmica mais extrema, chamada anafilaxia, os mediadores liberados por mastócitos bloqueiam vias aéreas até causar asfixia e causam um colapso cardiovascular que pode levar a morte. Os mastócitos possuem grânulos compostos de histamina, heparina e, em alguns casos, sulfato de condroitina (estes dois últimos são polissacarídeos sulfatados pertencentes à família dos glicosaminoglicanos) e proteases. Os proteoglicanos, incluindo a heparina e o sulfato de condroitina, são os principais constituintes tanto de mastócitos quanto de basófilos. São moléculas carregadas negativamente, assim servindo como matrizes de armazenamento de moléculas carregadas positivamente (como aminas biogênicas, proteases e outros mediadores). Os proteoglicanos liberam mediadores em diferentes taxas após a exocitose e, desta maneira, são capazes de controlar a cinética das reações de hipersensibilidade imediata. A desgranulação de mastócitos é considerada o componente central das doenças alérgicas e as manifestações clínicas e patológicas dependem dos tecidos onde os mediadores agem, bem como da cronicidade do processo inflamatório resultante (ABBAS, 2012).

Outras glicanas, além dos polissacarídeos contidos nos grânulos de mastócitos, estão envolvidas em processos alérgicos. As N-glicosilações de glicoproteínas de plantas e insetos são os determinantes imunológicos ambientais mais abundantes. Os principais motivos presentes nestes

organismos são xilose e “core” fucose em ligação α -1,3. As plantas possuem estes dois epitopos, enquanto insetos possuem apenas a fucose. Cerca de 20% dos pacientes alérgicos a plantas e insetos geram IgE anti-glicana, que frequentemente é acompanhada por IgG (ALTMANN, 2007).

Outro componente importante em algumas alergias é o receptor de manose. Trata-se de uma glicoproteína transmembrana tipo lectina tipo-C (*C-type lectin-like*) expressa pelas células dendríticas, que são células apresentadoras de antígenos e que reconhecem vários glicolárgenos de diversas fontes e estão envolvidos em promover resposta alérgica à poeira doméstica *in vitro*. Este receptor possui duas atividades lectinas, envolvendo reconhecimento de carboidratos cálcio-dependente através dos domínios tipo lectina tipo-C (*C-type lectin-like domains*, CTLDs), bem como ligação de glicanas sulfatadas acídicas independente de cálcio nas posições 3 ou 4 através do domínio rico em cisteína (*cysteine rich*, CR) (EMARA ET AL., 2011).

1.5 FEL D 1, O PRINCIPAL ALÉRGENO DO GATO

O gato doméstico (*Felis domesticus*) é um dos animais de estimação mais comuns em comunidades ocidentais, e seu pelo é uma fonte significativa de alérgeno (KROLL-KRISTENSEN ET AL., 1997), que contém diversas moléculas, sendo a proteína Fel d 1 o componente alérgeno majoritário (LEITERMANN E OHMAN, 1984). Esta proteína está presente em diversos tecidos do animal, como glândulas salivares, glândulas sebáceas, pele, pelo e cérebro (Figura 6) (SCHOU, 1993). O impacto destes alérgenos na saúde pública pode ser ilustrado no fato de que 20 a 30% dos indivíduos que sofrem de asma alérgica possuem alergia a gatos (GRÖNLUND ET AL., 2010).

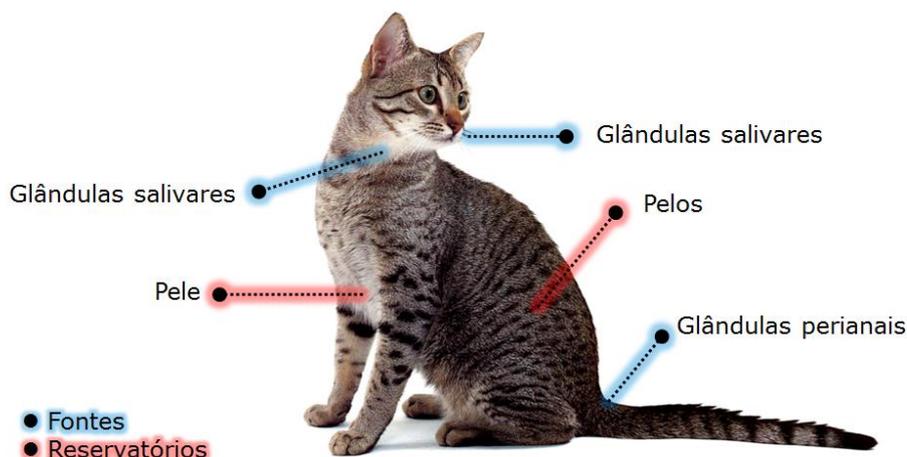


Figura 6. Representação esquemática da distribuição de Fel d 1 no gato doméstico, *Felis domesticus*.

A proteína Fel d 1 é um dímero de ~38 kDa não-covalentemente ligado composto de duas subunidades de ~18 kDa (também chamada tetrâmero por alguns autores, por ser um dímero de heterodímeros), pertencente à família das secretoglobinas. Cada subunidade (ou heterodímero) é composta por uma cadeia α e β orientadas de maneira antiparalela, unidas por três pontes dissulfeto (KROLL-KRISTENSEN ET AL., 1997). A orientação correta das cadeias α e β é considerada crítica, pois a atigenicidade relacionada a IgE é conformacionalmente dependente (DUFFORT ET AL., 1991; KROLL-KRISTENSEN ET AL., 1997; VAILES ET AL., 1994). Cada uma das subunidades de Fel d 1 contém uma glicana N-ligada à Asn33. Estudos de espectrometria de massas detectaram que sua estrutura é do tipo complexo, tri-ramificada e heterogênea (KROLL-KRISTENSEN ET AL., 1997), sendo a estrutura mínima a glicana [(-Asn)-GlcNAc₂Man₃GlcNAc₃Gal₂] e a máxima a glicana [(-Asn)-GlcNAc₂Man₃GlcNAc₃Gal₃(Fuc1)-NeuAc₃].

Uma estrutura cristalográfica de um tetrâmero recombinante de Fel d 1, expresso em *E. coli*, está disponível (PDB ID 2ejn). Nesta estrutura, foram revelados dois diferentes sítios de ligação a cálcio. Os sítios externos estão posicionados simetricamente em cada lado do tetrâmero (KAISER ET AL., 2007). Estes sítios correspondem a sítios putativos de ligação a Ca²⁺ sugeridos para a uteroglobina, uma molécula estruturalmente semelhante à Fel d 1, também da família das secretoglobinas, tipo citocina, induzida por esteroides com atividade

anti-inflamatória e propriedades imunomodulatórias (BARNES ET AL., 1996). A ligação a Ca^{2+} na uteroglobina é um assunto controverso, com grupos demonstrando esta ligação e outros demonstrando que esta proteína não se liga a Ca^{2+} (por isso os sítios são ditos “putativos”). O segundo sítio de ligação a Ca^{2+} no cristal de Fel d 1 encontra-se na interface de dimerização (Figura 7), estabilizando a formação do tetrâmero de Fel d 1 e induzindo mudanças conformacionais locais importantes que governam o formato de duas cavidades, uma em cada monômero, de tamanhos diferentes, preenchidas por água. Esta estrutura sugere um potencial portal para um ligante desconhecido. A presença dos íons Ca^{2+} levou à especulação de que a Fel d 1 poderia provocar uma resposta alérgica através da modulação da fosfolipase A2, por sequestrar íons Ca^{2+} , de maneira similar ao que foi proposto para a uteroglobina (ANDERSSON ET AL., 1994; NORD ET AL., 1995).

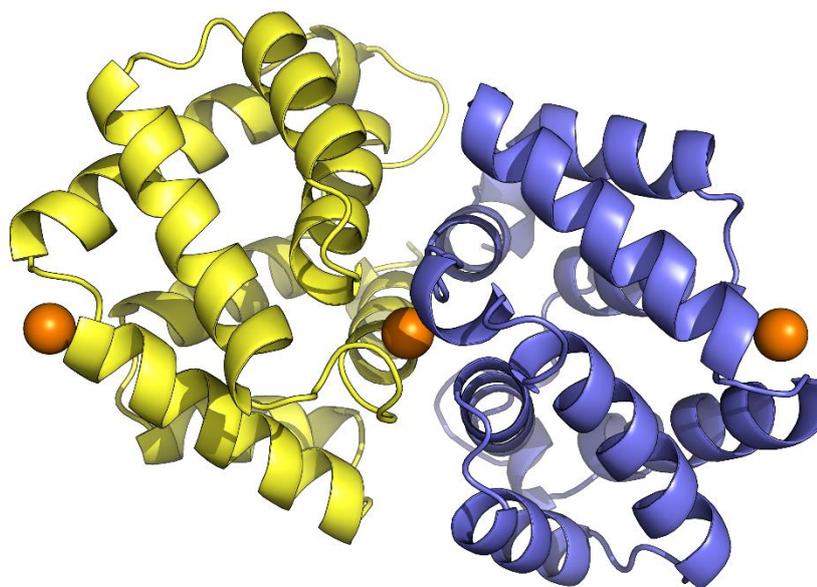


Figura 7. Estrutura do tetrâmero de Fel d 1 com os sítios de ligação a Ca^{2+} , conforme o PDB ID 2ejn. (Subunidades heterodiméricas representadas pelas cores amarela e azul e íons Ca^{2+} representados por esferas laranjas).

Alguns estudos sugerem importantes papéis estruturais e imunobiológicos da N-glicosilação na Fel d 1 (KROLL-KRISTENSEN ET AL., 1997; VAILES ET AL., 1994), enquanto outros estudos demonstraram que diferentes padrões de glicosilação não afetam a produção de IgE *in vitro* (SEPPÄLÄ ET AL., 2005). Um estudo mais recente *in vivo* demonstrou que o receptor de manose (glicoproteína transmembrana tipo lectina tipo-C expressa pelas células dendríticas, conforme

explicitado no item anterior) tem um papel essencial na internalização da proteína Fel d 1. O reconhecimento de Fel d 1 pelo receptor de manose é mediado pelo domínio CR (*cysteine rich*, rico em cisteína) deste receptor, o que está correlacionado com a presença de carboidratos sulfatados na Fel d 1 natural. Camundongos selvagens e “*knockout*” para o receptor de manose foram usados em um estudo, comparando a sensibilização à Fel d 1 entre estes dois grupos. Foi demonstrado que os camundongos sem o receptor produziram níveis menores de IgE e IgG em relação ao grupo selvagem, evidenciando que o receptor de manose tem um papel importante na sensibilização a este alérgeno *in vivo* (EMARA ET AL., 2011).

1.6 GLICANAS E CÂNCER

Alterações em padrões de glicosilação são uma característica de fenótipos tumorais. Esse fenômeno foi descrito pela primeira vez por Meezan e colaboradores em 1969, com a demonstração de que fibroblastos saudáveis tem menos glicoproteínas de membrana do que aqueles transformados (MEEZAN ET AL., 1969). Este fato foi mais tarde corroborado por evidências histológicas indicando que lectinas apresentam ligação a tecidos saudáveis diferente daquela a tecidos malignos (TURNER, 1992; SAUSSEZ ET AL., 1998). Mais recentemente, glicanas de superfície celular associadas a câncer foram diretamente caracterizadas usando anticorpos monoclonais específicos e espectrometria de massas (SHRIVER ET AL., 2004), elucidando as alterações moleculares que ocorrem em uma transformação maligna no câncer de mama (PANCINO ET AL., 1991) e colorretal (MATSUSHITA ET AL., 1990).

Diversas glicanas, tanto na superfície tumoral quanto em outros elementos do hospedeiro, têm sido identificadas como mediadores chave em eventos fisiopatológicos durante a progressão tumoral, que envolve várias alterações na sinalização intra e intercelular. Isso promove uma série de eventos, como: desregulação do ciclo celular, facilitando a proliferação; surgimento de células invasivas que se dissociam do tumor e migram para a matriz extracelular; convocação da rede endotelial neovascular próxima (angiogênese); promoção da disseminação de células tumorais com características de superfície celular,

que promovem interações de adesão com plaquetas, leucócitos e células do endotélio vascular sanguíneas e linfáticas; facilitação da evasão da imunidade inata. Finalmente, um conjunto de células “sobrevivente” invade, neovasculariza, dissemina, extravasa e prolifera em uma nova localização tecidual, tornando-se uma metástase (FUSTER E ESKO, 2005).

A maioria das classes de glicanas existe como glicoconjugados ligados a membranas ou como moléculas secretadas, que podem integrar a matriz extracelular. Estas localizações posicionam as glicanas para mediar adesão e mobilidade celular, bem como sinalização intracelular (MARTH, 1999). No ambiente tumoral, alterações de glicosilação permitem que as células neoplásicas assumam muitos dos eventos que ocorrem no desenvolvimento (como ativação de receptores, mobilidade e adesão celular), permitindo que as células tumorais invadam ou espalhem-se no organismo (VARKI, 1999). As transformações malignas são frequentemente acompanhadas por expressão de genes oncofetais (epitopos que são expressos em tecidos embrionários e células tumorais). Muitos dos primeiros anticorpos identificados específicos para tumores eram direcionados contra antígenos oncofetais sacarídicos presentes em glicoproteínas e glicoesfingolipídeos tumorais. Em alguns casos, a subexpressão e alteração ou encurtamento na ramificação de certas glicanas está relacionado ao crescimento celular. Alterações similares em tumores permitem que eles intensifiquem sua capacidade proliferativa (VARKI, 1999; FEIZI, 1985).

1.7 GLICOSILTRANSFERASES

As GTs estabelecem ligações glicosídicas, realizando a biossíntese de dissacarídeos, oligossacarídeos ou polissacarídeos ao catalisarem a transferência de um monossacarídeo, a partir de moléculas doadoras, para uma molécula aceptora (TANIGUCHI ET AL., 2002). Este grupo de enzimas apresenta especificidades variadas por diferentes substratos, assim como localização celular diversa, sendo portanto responsáveis pela diversidade e complexidade de glicoconjugados, glicoproteínas, glicolipídeos e proteoglicanas.

Os mamíferos utilizam apenas 9 tipos de doadores nucleotídeo-sacarídicos para GTs: UDP-glicose, UDP-galactose, UDP-GlcNAc, UDP-GalNAc, UDP-xilose, UDP-ácido glicurônico, GDP-manose, GDP-fucose e CMP-ácido siálico. Os grupos fosfato destas moléculas doadoras são coordenados às enzimas por cátions divalentes, mas existem enzimas independentes de metais.

As GTs podem ser divididas em enzimas de “retenção” ou “inversão” de acordo com a estereoquímica da ligação do carbono anomérico do doador (retenção ($\alpha \rightarrow \alpha$) ou inversão ($\alpha \rightarrow \beta$) da estereoquímica durante a transferência). O mecanismo de inversão requer apenas um ataque nucleofílico a partir do átomo aceptor para inverter a estereoquímica (nesse caso a própria hidroxila aceptor é quem realiza o ataque nucleofílico, aproximando-se do carbono anomérico pelo lado oposto). O mecanismo de retenção ainda está sendo debatido, mas há evidências contrárias ao mecanismo de deslocamento duplo (que causaria duas inversões) e ao mecanismo dissociativo, e trabalhos recentes sugerem fortemente um mecanismo de ataque ortogonal, o que estaria de acordo com as estruturas disponíveis, geometria e estudos cinéticos e computacionais (**Figura 8**) (SCHUMAN ET AL., 2013).

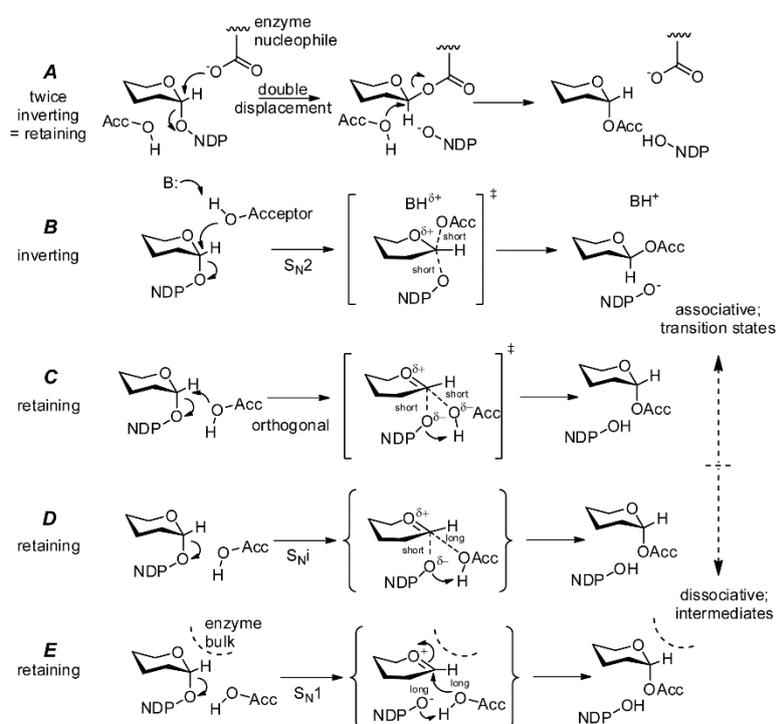


Figura 8. A) Mecanismo de deslocamento duplo, envolvendo um intermediário covalente glicosil-enzima. B) Inversão por S_N2 , com ataque em C1 pelo lado oposto pela hidroxila aceptor. Mecanismos alternativos para retenção: C) Mecanismo ortogonal, com ataque nucleofílico em C1 pelo aceptor; D) Mecanismo S_Ni ; E) Mecanismo S_N1 (SCHUMAN ET AL., 2013).

O baixo grau de identidade de sequências entre as diversas GTs, acompanhado por um reduzido número de estruturas cristalinas, vem limitando nossa capacidade de estudar os aspectos estruturais desta família de enzimas. Entretanto, determinações estruturais nos últimos anos revelaram que a maioria das glicosiltransferases apresenta um de dois tipos de enovelamento, designados GT-A e GT-B (BOURNE Y E HENRISSAT B, 2001; COUTINHO ET AL., 2003).

O enovelamento GT-A consiste de um sanduíche $\alpha/\beta/\alpha$ similar a um enovelamento do tipo Rossmann (Figura 9A). Uma característica geral desta família é a presença de um motivo comum DxD e a necessidade de um cátion divalente para a atividade (BRETON ET AL., 1998; BRETON & IMBERTY, 1999). Uma comparação entre domínios catalíticos das enzimas desta família revelou duas regiões estruturalmente conservadas em todos os membros, incluindo enzimas que agem por retenção e inversão e, assim, sugerem que elementos estruturais comuns são necessários para a transferência glicosídica, independente da estereoquímica da reação.

O enovelamento GT-B consiste de dois enovelamentos Rossmann com uma região conectora e um sítio catalítico localizado entre os domínios (Figura 9B). Há uma ótima conservação estrutural entre membros desta família, especialmente no domínio C-terminal, onde ocorre a ligação ao substrato. As maiores variações ocorrem no domínio N-terminal, nas alças e hélices que apontam para o sítio catalítico e que evoluíram para acomodar diferentes aceptores (BRETON ET AL., 2006).

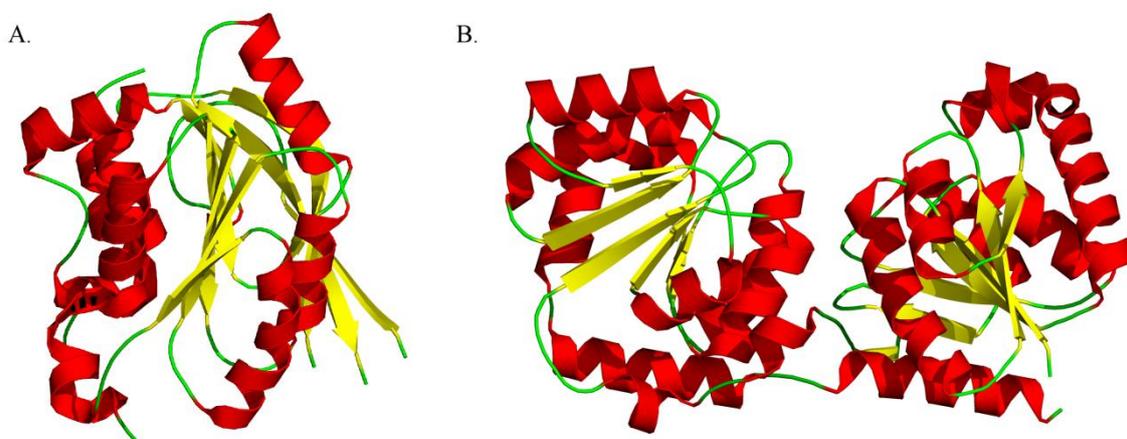


Figura 9. A) Enovelamento GT-A (PDB ID 1h7l) e B) Enovelamento GT-B (PDB ID 1nlm).

O substrato aceptor de uma glicosiltransferase pode ser simples, como uma unidade monossacarídica, ou complexo, como oligo e polissacarídeos, ácidos nucleicos, lipídeos ou proteínas. A transferência glicosídica no caso de proteínas, pode ocorrer principalmente em resíduos de tirosina, serina ou treonina (gerando glicoproteínas O-ligadas) ou então em resíduos de asparagina (gerando glicoproteínas N-ligadas) (Figura 10).

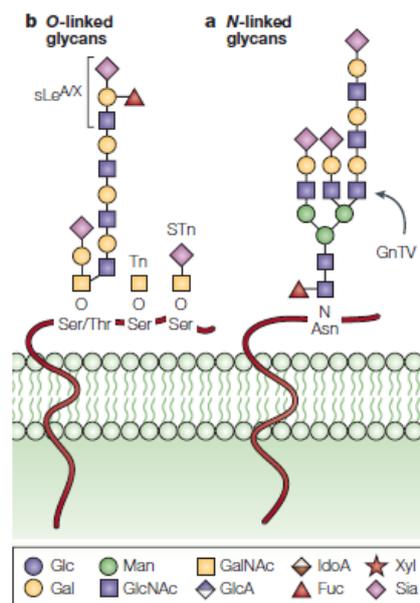


Figura 10. A) N-ligadas e B) O-ligadas. (Adaptado de FUSTER & ESKO, 2005).

Além destes dois tipos principais de glicosilação, existem outros menos comuns, descritos mais recentemente: fosfoglicosilação, C-glicosilação (presentes apenas em eucariotos) e gliptação (identificada em eucariotos e arqueas). Na fosfoglicosilação os oligossacarídeos estão ligados a resíduos de serina ou treonina via ligações fosfodiéster (SPIRO, 2002). Na C-manosilação ou C-glicosilação um resíduo de α -manose é adicionado ao grupamento indol de um triptofano presente na primeira posição da sequência consenso Trp-X1-X2-Trp/Cys, sendo que essa ligação é favorecida quando X1 é um aminoácido pequeno ou polar (como serina, alanina, glicina ou treonina) e desfavorecida quando X1 é fenilalanina ou leucina (HOFSTEENGE ET AL., 1994; DE BEER ET AL., 1995; KRIEG ET AL., 1997; DOUCEY ET AL., 1998; KRIEG ET AL., 1998; JULENIUS, 2007). E na gliptação ocorre a adição de glicofosfatidilinositol (GPI) à região C-

terminal de polipeptídeos, que é realizada a fim de ancorar algumas glicoproteínas e proteoglicanas a superfícies celulares (LOW, 1989; SPIRO, 2002; PAULICK E BERTOZZI, 2008).

1.7.1 FUCOSILTRANSFERASE 8

A fucosiltransferase 8 (FUT8) é uma α -1,6 fucosiltransferase que catalisa a transferência de uma Fuc (a partir de uma GDP- β -L-fucose) para a primeira GlcNAc de um oligossacarídeo N-ligado (“*core fucosylation*”) (Figura 11). Este tipo de fucosilação é frequentemente observado em N-glicanas de α -fetoproteína, um marcador tumoral para carcinoma hepatocelular bastante conhecido (TAKETA ET AL., 1993; NODA ET AL., 1998).

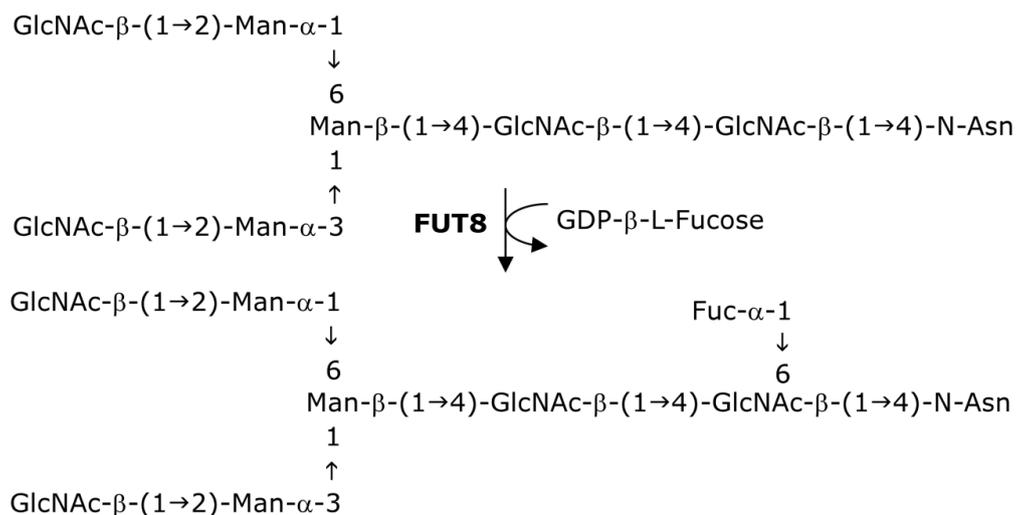


Figura 11. Reação de α 1,6 fucosilação catalisada pela FUT8 (transferência de uma fucose, em ligação α 1,6, para a primeira GlcNAc de uma glicana N-ligada).

Experimentos com FUT8 de mamíferos demonstraram que a metástase é suprimida pela fucosilação α -1,6 da integrina α 5 β 1 (MIYOSHI ET AL., 1999). Já a fucosilação α -1,6 das N-glicanas de imunoglobulina humana IgG1 regula a citotoxicidade celular anticorpo-dependente (ADCC). A ausência de “*core*” fucose em IgG1 resulta em um aumento de atividade de ADCC em até 100 vezes, sugerindo que IgG1 deficiente nesta fucose seria útil no tratamento de câncer utilizando terapia com anticorpo (SHIELDS ET AL., 2002; SHINKAWA ET AL., 2003). Além destes experimentos, foi verificado que a fucosilação α -1,6 regula

a função de imunoglobulinas por modificar suas características físico-químicas (OKAZAKI ET AL., 2004). Outro grupo também reportou que o rompimento do gene FUT8 em camundongos levou a fenótipos de retardamento de crescimento, enfisema pulmonar e até mesmo morte (WANG ET AL., 2005). Estes fenótipos ocorreram principalmente pela falta de fucosilação α -1,6 em fator de crescimento epidermal, receptores de fator de crescimento β (WANG ET AL., 2005; TANIGUCHI ET AL., 2006) e outras moléculas (LEE ET AL., 2006; LI ET AL., 2006; ZHAO ET AL., 2006). Todos estes estudos sugerem fortemente que FUT8 e seu produto enzimático (“core” fucose) tem papéis em vários eventos fisiológicos e fisiopatológicos.

A estrutura de FUT8 humana foi determinada por cristalografia de raios X (PDB ID 2de0), correspondente aos resíduos 68 a 575 (regiões haste e transmembrana ausentes) (IHARA ET AL., 2007). Foi observada a presença de 3 domínios: um domínio N-terminal (resíduos 109 a 173) contendo duas longas α -hélices antiparalelas formando uma estrutura “*coiled-coil*”; um domínio catalítico (resíduos 174 a 501) composto por uma estrutura de folha aberta α/β (resíduos 203 a 297) e um enovelamento Rossmann (resíduos 359 a 492), ligadas por três hélices; e um domínio SH3 C-terminal (resíduos 502 a 562) (Figura 12). Nesta estrutura cristalográfica os resíduos Asp368 a Thr372 estão desordenados.

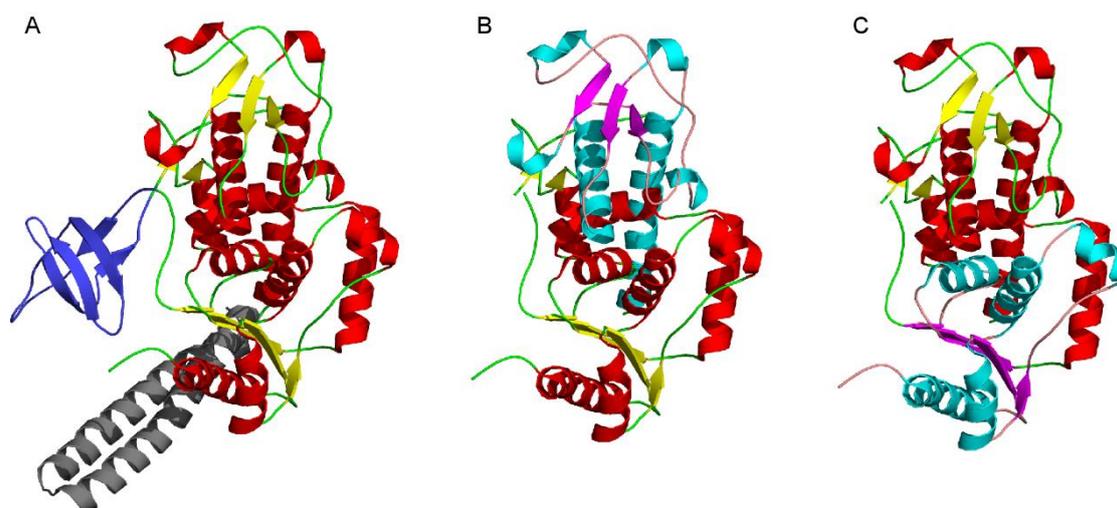


Figura 12. A) Estrutura cristalográfica de FUT8 humana (PDB ID 2de0). Domínio N-terminal (*coiled-coil*) em cinza. Domínio catalítico colorido de acordo com a estrutura secundária (α -hélices em vermelho, folhas- β em amarelo e alças em verde). E domínio C-terminal SH3 em azul. B) Domínio catalítico, destacando a estrutura de sanduíche $\alpha/\beta/\alpha$ (α -hélices em azul ciano, folhas- β em roxo/rosa e alças em salmão). C) Domínio catalítico, destacando a estrutura tipo Rossmann (α -hélices em azul ciano, folhas- β em roxo/rosa e alças em salmão).

Estudos comparativos utilizando a ferramenta DALI (HOLM E SANDER, 1993) demonstraram que, quando analisados separadamente, os domínios de FUT8 apresentam similaridade estrutural com outras proteínas (IHARA ET AL., 2007). Entretanto, a estrutura completa de FUT8 apresentou baixa similaridade estrutural com proteínas conhecidas (IHARA ET AL., 2007). A estrutura da FUT8 apresenta um enovelamento Rossmann, similar a GTs da família GT-A. Entretanto, pesquisas usando o domínio catalítico (resíduos 201 a 500) resultaram em similaridades estruturais com enzimas da família GT-B. Além disso, a FUT8 não possui um motivo DxD e possui atividade total na ausência de íon metálico. Esses resultados dão suporte para classificar esta enzima como semelhante à família GT-B.

1.7.2 CORE-2 N-ACETIL GLICOSAMINILTRANSFERASE LEUCOCITÁRIA

A Core-2 N-acetil glicosaminiltransferase leucocitária (C2GnT-L) é uma GT presente no Golgi *cis/medial* que catalisa a conversão de O-glicanas de estrutura *Core-1* para estruturas *Core-2* (WILLIAMS ET AL., 1980; BIERHUIZEN E FUKUDA, 1992). A C2GnT-L é uma N-acetil glicosaminiltransferase que age por inversão (COUTINHO ET AL., 2003), transferindo uma GlcNAc, em ligação 1,6, de uma UDP-GlcNAc para Gal- β (1-3)-GalNAc-O-Ser/Thr ("*Core*" 1) para gerar Gal- β (1-3)[GlcNAc- β (1-6)]GalNAc-O-Ser/Thr ("*Core*" 2) (Figura 13). Íons divalentes, essenciais para a atividade catalítica em muitas GTs, não são necessários em C2GnT-L (YEH ET AL., 1999; SCHWIENTEK ET AL., 2000).

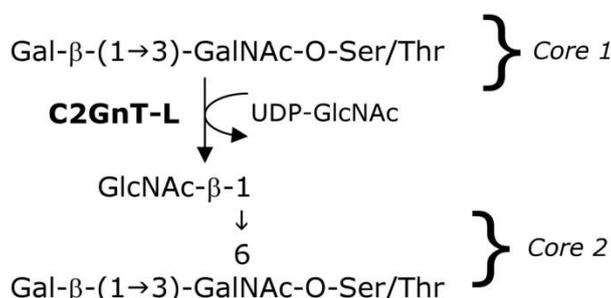


Figura 13. Reação catalisada por C2GnT-L (transferência de GlcNAc, em ligação 1,6, para uma estrutura *Core* 1, formando a estrutura *Core* 2).

As O-glicanas *Core-2* tem demonstrado importante participação na migração de linfócitos e rolamento de leucócitos mediados por selectinas. L-selectinas de linfócitos ligam-se a O-glicanas de células endoteliais contendo 6-sulfo sialil Lewis X em estruturas *Core-2* e *Core-1* estendidas, enquanto neutrófilos que expressam sialil Lewis X em O-glicanas *Core-2* ligam-se a E- e P-selectinas em células endoteliais (KAWASHIMA ET AL., 2005; SPERANDIO ET AL., 2001; LOWE, 2003).

A expressão de C2GnT-L está correlacionada com a progressão tumoral de vários tipos de câncer, tornando esta enzima extremamente interessante no estudo de metástase tumoral. Ela é superexpressa em câncer de pulmão, próstata e colorretal e alguns trabalhos demonstraram que a transfecção de C2GnT-L para linhagens celulares de câncer de próstata levou ao aumento do tamanho do tumor (MACHIDA ET AL., 2001; HAGISAWA ET AL., 2005).

A estrutura de C2GnT-L de murino foi determinada por cristalografia de raios X (PDBIDs 2gak e 2gam) (PAK ET AL., 2006). Esta estrutura compreende os resíduos 38 a 428 (ausentes os resíduos 1 a 16 do domínio N-terminal citoplasmático e os resíduos transmembranares 17 a 34). A primeira região (resíduos 38 a 121) é composta apenas de α -hélices e corresponde a uma haste que estende o domínio catalítico ao lúmen do Golgi. A segunda região corresponde ao domínio catalítico (resíduos 122 a 428) e é uma estrutura $\alpha/\beta/\alpha$ contendo 6 folhas- β flanqueadas em um lado por α -hélices e 3 folhas- β e no outro lado por α -hélices e 2 e 3 folhas- β (Figura 14).

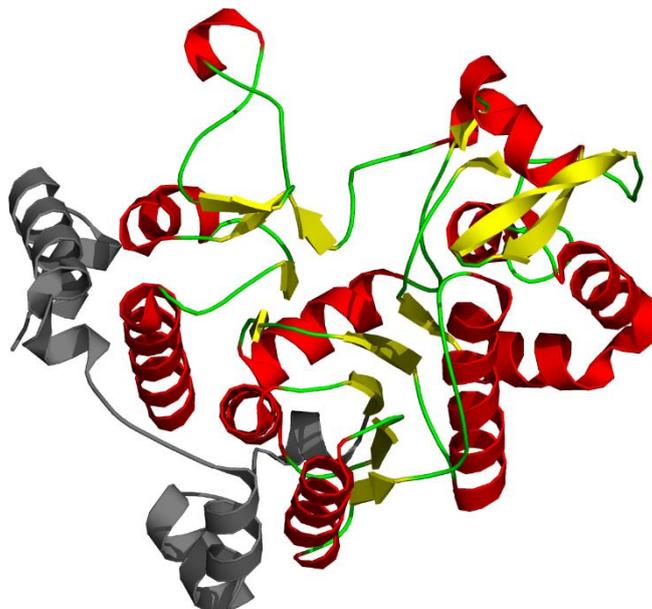


Figura 14. Estrutura cristalográfica de C2GnT-L (PDB ID 2gak). Resíduos 38 a 121 (α -hélices) coloridos em cinza. O restante corresponde ao domínio catalítico (resíduos 122 a 428), coloridos de acordo com a estrutura secundária (α -hélices em vermelho, folhas- β em amarelo e alças em verde).

A estrutura do domínio catalítico pertence ao enovelamento GT-A. Os resíduos 122 a 240 contêm o sítio de ligação para a porção nucleotídica do substrato doador (resíduos correspondentes àqueles conservados entre as GTs da família GT-A) (TARBOURIECH ET AL., 2001), enquanto os resíduos 241 a 428 contêm o sítio de ligação do substrato aceptor.

Apesar da estrutura de C2GnT-L possuir enovelamento tipo GT-A, o motivo característico de ligação a íon metálico DxD está ausente. Alinhamentos estruturais sugerem que o papel dos íons Mn^{2+} ou Mg^{2+} em estruturas GT-A dependentes de metais é suprido pelos resíduos de aminoácidos básicos Arg378 e Lys401 em C2GnT-L. O uso de aminoácidos básicos é similar ao comportamento observado em GTs tipo GT-B e fornece evidências estruturais de convergências do mecanismo catalítico entre as GTs de enovelamentos GT-A e GT-B (PAK ET AL., 2006).

A C2GnT-L apresenta dois sítios de N-glicosilação, nos resíduos Asn58 e Asn95 (ambos na região N-terminal). Estudos verificaram a ocupação e importância destes sítios, sendo que a glicosilação em Asn95 é essencial para a atividade biológica de C2GnT-L - gerar ligantes funcionais de selectinas,

especialmente PSGL-1 (*P-selectin glycoprotein ligand-1*), bem como para localizá-la no Golgi *cis/medial* (HAMON ET AL., 2005).

1.8 DINÂMICA MOLECULAR

O uso de simulações de dinâmica molecular (DM), utilizado para biomoléculas desde a década de 70 (MCCAMMON ET AL., 1977), é hoje uma ferramenta computacional amplamente difundida, utilizada para investigação da estrutura e dinâmica de biomoléculas em geral, abrangendo estudos de interações entre compostos e suas proteínas-alvo, desnaturação e enovelamento proteicos (PONDER E CASE, 2003), entre outros. Na DM os átomos são tratados como partículas se movendo sob a influência da mecânica clássica (MAGINN E ELLIOT, 2010) e as sucessivas configurações do sistema são geradas pela integração da equação do movimento de Newton:

$$\frac{d^2r_i}{dt^2} = \frac{F_i}{m_i}$$

Nesta equação, d^2r_i/dt^2 é a aceleração, m_i é a massa e F_i é a força sobre um determinado átomo i . O resultado desta integração empregada sobre todos os átomos do sistema é uma trajetória que especifica como as posições e velocidades das partículas do sistema variam com o tempo relativo à aceleração (LEACH, 2001). Entretanto, esta equação não consegue determinar a magnitude (módulo) e a direção da força F_i sobre os átomos do sistema, nem sua relação com as características de cada molécula em estudo. Estes parâmetros serão calculados em função de mudanças de energia potencial (V) entre a posição atual e a posição seguinte (aquela que representará o próximo passo da simulação) sobre cada átomo (coordenadas atômicas r_i a r_n):

$$F_i = \frac{-\partial V(r_i, \dots, r_n)}{\partial r_i}$$

A função de energia potencial (V_{total}) é dada conforme a [Figura 15](#), sendo descrita no chamado Campo de Força empregado na simulação. O campo de força pode ser definido como um conjunto de funções e parametrizações usadas em cálculos de mecânica molecular (DE SANT'ANNA, 2002) que definem as energias associadas a interações ligadas e não-ligadas. Por exemplo, entre

átomos ligados covalentemente o estiramento de ligação e distorção do ângulo de ligação (tanto de valência quanto de diedro) de uma molécula é descrito em relação à sua conformação não tensionada (caracterizada pelos valores-padrão de comprimentos e de ângulos de ligação). Juntamente a termos dedicados à descrição da interação entre átomos não-ligados, como forças eletrostáticas, ligação de hidrogênio e forças de van der Waals, dentre outros, os campos de força expressam o somatório das funções de energia potencial de cada átomo e calculam a energia dos sistemas em função das posições dos núcleos dos átomos (LEACH, 2001).

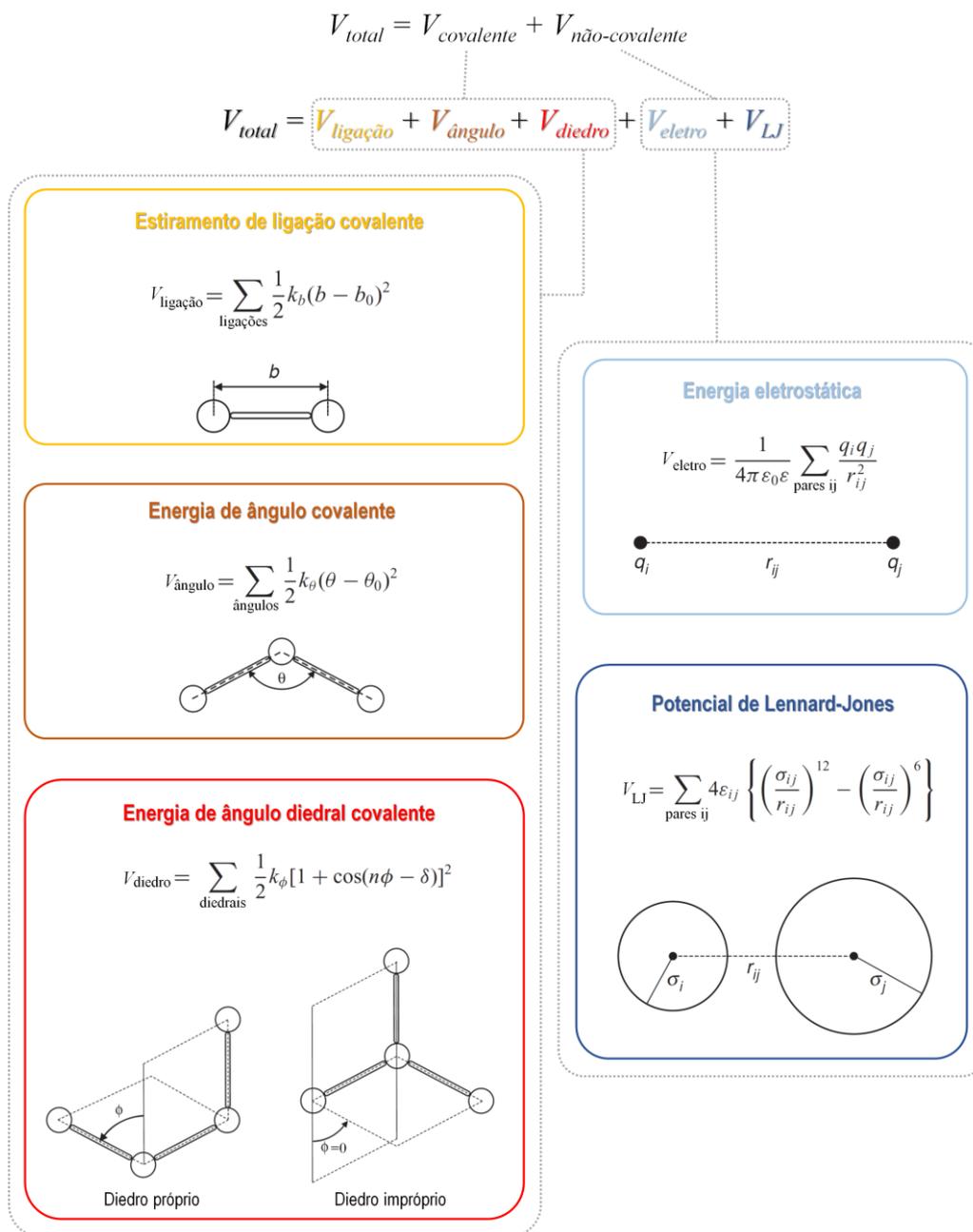


Figura 15. Fun\~{c}oes de energia que compoem campos de for\~{c}a empregados em din\~{a}mica molecular (LIGABUE-BRAUN, 2014).

Existem diversos campos de for\~{c}a dispon\~{i}veis para simula\~{c}oes de DM. No pacote de simula\~{c}oes GROMACS (VAN DER SPOEL ET AL., 2005) destacam-se o AMBER (CASE ET AL., 2005), o CHARMM (MACKERELL ET AL., 1998), e o GROMOS (VAN GUNSTEREN ET AL., 1996). Destes, AMBER e CHARMM s\~{a}o campos chamados “*all-atom*”, ou seja, apresentam par\~{a}metros para todos os \acute{a}tomos presentes no sistema, enquanto o GROMOS \acute{e} um campo “*united-atom*”,

onde não são incluídas representações para hidrogênios não polares, apenas hidrogênios polares estão incluídos nas definições do campo.

2. OBJETIVOS

A partir dos itens apresentados na introdução, percebe-se que que glicanas, glicoproteínas e glicosiltransferases possuem papéis em diversos processos fisiopatológicos. Desta maneira, o objetivo geral deste trabalho é a caracterização estrutural de três proteínas – Fel d 1, FUT8 e C2GnT-L -, através do refinamento de suas estruturas cristalográficas, introduzindo informação sacarídica.

Considerando o papel da glicosilação e dos sítios ligadores de Ca^{2+} nos processos alérgicos, o presente trabalho pretende estudar o principal alérgeno do gato, a proteína Fel d 1, através dos seguintes objetivos específicos:

- ▶ Verificação do papel estrutural dos íons Ca^{2+} nos sítios propostos para Fel d 1 através de cálculos semiempíricos e moleculares em solução;
- ▶ Caracterização da proteína Fel d 1 em solução na forma glicosilada, em diferentes glicoformas (correspondentes às formas mínima e máxima de estruturas sacarídicas).

Além disso, considerando a superexpressão das glicosiltransferases FUT8 e C2GnT-L em diversos tipos tumorais, bem como a ausência de estudos estruturais e conformacionais destas moléculas em solução, o presente trabalho visa também ampliar o conhecimento acerca dessas enzimas, aplicando técnicas computacionais ao estudo de diferentes aspectos de sua biologia estrutural. Neste contexto foram traçados os seguintes objetivos específicos:

- ▶ Caracterização estrutural e conformacional da FUT8 livre em solução e complexada ao seu substrato doador;
- ▶ Caracterização da interação da FUT8 com sua glicana-alvo e com uma proteína alvo (Integrina $\alpha 5\beta 1$);
- ▶ Caracterização da glicobiologia estrutura de C2GnT-L em solução, livre e complexada a seus substratos doador e aceptor;

Espera-se, a partir dos resultados alcançados, contribuir na compreensão do funcionamento destas proteínas, possibilitando futuros estudos acerca de seus papéis em enfermidades.

3. MÉTODOS

3.1 SISTEMAS ESTUDADOS

Neste trabalho foram submetidos à DM, sob solvente explícito, os seguintes sistemas:

- ▶ Fel d 1 tetramérica, na ausência e na presença de íons Ca^{2+} , em três campos de força diferentes (Amber, Charmm e Gromos), totalizando 6 sistemas, $n=1$ para cada um;
- ▶ Fel d 1 tetramérica na presença de íons Ca^{2+} contendo a forma mínima de glicosilação, $n=1$;
- ▶ Fel d 1 tetramérica na presença de íons Ca^{2+} contendo a forma máxima de glicosilação, $n=1$;
- ▶ Unidade cristalina da Fel d 1, constituída de dez tetrâmeros adicionais à unidade tetramérica central, $n=1$;
- ▶ FUT8 não complexada sem a alça DKVGT (resíduos 368-372, desordenados no cristal) nos campos de força Gromos ($n=2$) e Amber ($n=2$);
- ▶ FUT8 complexada ao substrato (UDP- β -L-glicose) sem a alça DKVGT ($n=3$);
- ▶ FUT8 não complexada com a alça DKVGT ($n=2$);
- ▶ FUT8 complexada ao substrato (UDP- β -L-glicose) com a alça DKVGT ($n=2$);
- ▶ FUT8 complexada à glicana alvo (DM) ($n=2$);
- ▶ C2GnT-L livre em solução ($n=2$, a partir de 2 estruturas cristalinas diferentes);
- ▶ C2GnT-I complexada ao seu substrato acceptor (Gal- β -(1-3)-GalNAc) ($n=1$);
- ▶ DM de C2GnT-L complexada ao seu substrato doador (UDP-GlcNAc) ($n=1$).

Adicionalmente, foi realizado cálculo semiempírico para Fel d 1 tetramérica, não glicosilada, com os quatro íons Ca^{2+} encontrados na estrutura cristalográfica.

Foi também realizado o atracamento molecular da FUT8 à sua glicana alvo, bem como o atracamento molecular da FUT8 à Intergina $\alpha 5\beta 1$.

3.2 PROGRAMAS UTILIZADOS

Durante a realização deste trabalho os seguintes programas foram utilizados:

- ▶ Pacote de simulações GROMACS 4.5 para a DM (HESS ET AL., 2008);
- ▶ Programa de cálculo semiempírico MOPAC (STEWART, 2008);
- ▶ Programa de modelagem comparativa MODELLER (FISER ET AL., 2000);
- ▶ Autodock 4.2 para atracamentos ligantes-proteína (MORRIS ET AL., 2009);
- ▶ ClusPro 2.0 para atracamento proteína-proteína (COMEAU ET AL., 2004);
- ▶ Programas de visualização de moléculas: VMD (HUMPHREY ET AL., 1996), Pymol (DELANO, 2002) e Swiss-PdbViewer (GUEx E PEITSCH, 1997);

3.3 PROTOCOLOS

3.3.1 DM

O protocolo geral de simulação dos sistemas descritos no item 3.1 foi baseado em procedimentos previamente descritos (por exemplo, por DE GROOT E GRUBMÜLLER, 2001; SACHETT E VERLI, 2011; POL-FACHIN E VERLI, 2012). Para todas as proteínas foram realizados cálculos de DM com o pacote GROMACS 4.5 (HESS ET AL., 2008) empregando o campo de força GROMOS96 53a6 (OOSTENBRINK ET AL., 2004). Para Fel d 1 e FUT8 livre foram também realizadas simulações com o campo de força Amber99SB-ILDN (LINDORFF-LARSEN ET AL., 2010). E, para Fel d 1, foram também realizadas simulações com o campo de força CHARMM27 (MACKERELLET AL., 2001).

Os sistemas foram solvatados em caixas triclinicas, usando condições periódicas de contorno e os modelos de água recomendado para cada campo de força. Para o Gromos foi usado o modelo SPC, enquanto para o Amber e Charmm o modelo Tip3P foi empregado. Contra-íons (Na^+ ou Cl^-) foram adicionados para neutralização dos sistemas. O método Lincs (HESS ET AL., 1997) foi aplicado na restrição de ligações covalentes de forma a permitir um passo de integração de 2 fs, enquanto as interações eletrostáticas foram calculadas utilizando o método Particle-Mesh Ewald (DARDEN ET AL., 1993). A temperatura e a pressão de cada sistema foram mantidas constantes através do acoplamento das proteínas, íons e solvente a banhos externos de temperatura e pressão, utilizando constantes de acoplamento de, respectivamente, $\tau = 0,1$ ps e $\tau = 0,5$ ps (BERENDSEN ET AL., 1984). A constante dielétrica do meio foi tratada como $\epsilon = 1$.

As simulações por dinâmica molecular geralmente iniciam com uma etapa de equilíbrio. Nessa etapa ocorre o aquecimento gradativo do sistema, visando uniformizar as energias contidas na estrutura analisada (seja cristalográfica, derivada de RMN ou obtida por modelagem comparativa) para que sejam evitadas deformações nas moléculas em estudo. Para equilíbrio, após 1 ps de DM sob restrição de posição (1.000 kJ/mol), cada um dos sistemas foi aquecido lentamente de 50 K a 300 K de maneira que, em cada um dos seis passos de 5 ps, há o aumento da temperatura em 50 K. A simulação teve prosseguimento sem restrições na temperatura de equilíbrio de 300 K, por 500 ns, considerando um valor de referência de 3.5 Å entre átomos pesados para ligações de hidrogênio, e um ângulo de corte de 30° entre doadores e aceptores de hidrogênio (HESS ET AL., 2008).

Os sistemas com a proteína Fel d 1 foram simulados na ausência e presença dos íons Ca^{2+} , nos três campos de força mencionados, totalizando 6 sistemas. Ainda, foram simuladas duas glicofomas da proteína, uma delas contendo em cada cadeia (no resíduo Asn103) uma N-glicana completa [(-Asn-) $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_3\text{Gal}_3$ (Fuc1)-NeuAc₃] e outra forma contendo em cada cadeia uma glicana reduzida [(-Asn-) $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_3\text{Gal}_2$], no campo de força Gromos96 53a6, formando então 8 sistemas que foram simulados por 100ns. A construção das estruturas sacarídicas foi feita utilizando o servidor Glycosciences (www.glycosciences.de). As geometrias das ligações glicosídicas

foram ajustadas para as conformações majoritárias em solução de cada ligação, obtidas em estudos anteriores por DM dos dissacarídeos em água (FERNANDES ET AL., 2010; POL-FACHIN ET AL., 2012). Um 9º sistema, contendo uma mimetização do ambiente cristalino da Fel d 1 (adicionando 10 tetrâmeros), gerado com a ferramenta Symexp do programa Pymol (DELANO, 2002), foi também simulado nos três campos de força mencionados por 25ns.

Para FUT8, nos sistemas com o substrato GDP- β -L-fucose foram utilizados parâmetros de fosfatos e guanosina presentes no campo de força Gromos 53a6 (OOSTENBRINK ET AL., 2004), juntamente com parâmetros da fucose previamente descritos pelo nosso grupo de pesquisa (POL-FACHIN ET AL., 2012). A estrutura cristalográfica disponível para FUT8 apresenta a região dos resíduos Asp368 a Thr372 desordenada, cujos autores da cristalografia acreditam ser uma alça flexível, já que em muitas GTs alças flexíveis são importantes para o mecanismo catalítico (IHARA ET AL., 2007). Desta maneira, foram estudados sistemas de FUT8 livre em solução e complexada ao substrato na presença e ausência desta alça, a fim de verificar a importância da mesma. A alça DKVGT foi modelada utilizando o programa MODELLER (FISER ET AL., 2000).

As simulações por DM de C2GnT-L utilizaram as mesmas condições descritas para Fel d 1 e FUT8, mas neste caso foram realizadas apenas simulações com o campo de força Gromos96 53a6. No banco de dados *Protein Data Bank* (PDB) existiam duas estruturas disponíveis: PDB ID 2gak (C2GnT-L “apo”) e PDB ID 2gam (C2GnT-L complexada ao substrato acceptor, Gal- β -(1-3)-GalNAc). Foram estudadas ambas as estruturas, 2gak e 2gam (retirando-se o substrato desta última), a fim de comparar os cristais. Posteriormente, a estrutura completa do cristal 2gam foi estudada, complexada ao substrato acceptor.

Para estudar o complexo com o substrato doador (UGP-GlcNAc), foi realizado primeiramente a ancoragem molecular deste ligante ao receptor (C2GnT-L), através do programa Autodock 4.2 (MORRIS ET AL., 2009). Após, seguiu-se o protocolo de DM conforme os demais sistemas.

3.3.2 Cálculo semiempírico de Fel d 1 com íons Ca^{2+}

Para descrever a coordenação dos íons Ca^{2+} , a proteína Fel d 1 foi submetida a cálculo semiempírico com o programa MOPAC (STEWART, 2008) contendo os quatro íons Ca^{2+} encontrados na estrutura cristalográfica. O protocolo empregou o método PM6 (STEWART, 2007) e a técnica MOZYME (STEWART, 1996), que compreende os passos a seguir: otimização dos átomos de hidrogênio a um gradiente abaixo de 5 kJ/mol/Å; otimização global com critério de gradiente de convergência de 10 kJ/mol/Å e cutoff = 6; otimização global com critério de gradiente de convergência de 5 kJ/mol/Å e cutoff = 9. Os cálculos foram realizados na ausência de moléculas de águas cristalográficas, como o modelo de solvatação contínua COSMO (KLAMT E SCHÜÜRMAN, 1993), e com todas as moléculas de água cristalográfica a um raio de 5 Å dos íons Ca^{2+} . Foram avaliadas as ordens de ligação e as distâncias para estimar os possíveis sítios de coordenação Ca^{2+} -aminoácido.

3.3.3 Atracamento de FUT8 à glicana-alvo

A estrutura da glicana alvo da fucosilação foi construída no portal *Glycosciences* (www.glycosciences.de) utilizando a ferramenta SWEET2 (Figura 16), e posteriormente ajustou-se as geometrias das ligações glicosídicas para as conformações majoritárias em solução de cada ligação, obtidas em estudos anteriores por DM dos dissacarídeos em água (FERNANDES ET AL., 2010; POL-FACHIN ET AL., 2012). O atracamento molecular à glicana alvo da fucosilação foi realizado no Autodock 4.2 (MORRIS ET AL., 2009). Neste processo, o receptor (FUT8) foi mantido rígido e as ligações glicosídica do ligante (glicana) foram mantidas flexíveis.

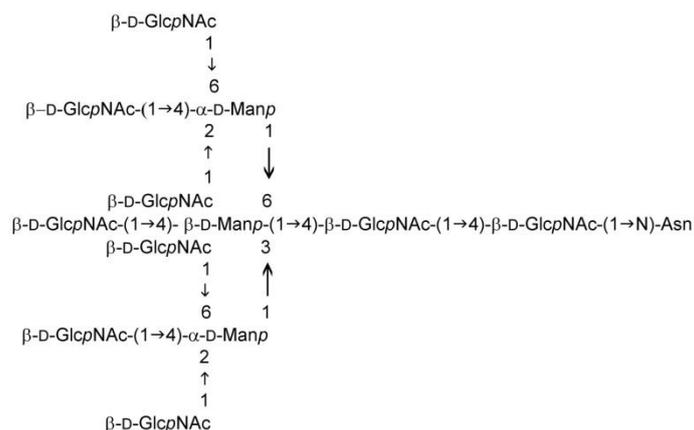


Figura 16. Estrutura da glicana alvo da fucosilação, posteriormente submetida ao atracamento molecular.

3.3.4 Atracamento de FUT8 à integrina $\alpha 5\beta 1$

Para o atracamento molecular da FUT8 à Integrina $\alpha 5\beta 1$ (PDB ID 3vi3) foi utilizado o programa PIPER (KOZAKOV ET AL., 2006), implementado no ClusPro 2.0 (COMEAU ET AL., 2004). Sabe-se que 50% das glicanas N-ligadas nesta Integrina são fucosiladas na primeira GlcNAc (“core” fucosilação) (LITYNSKA ET AL., 2002; NAKAGAWA ET AL., 1996). Desta maneira, foram selecionados os sítios de N-glicosilação, presentes nas cadeias A, B, C e D, para o atracamento.

4. RESULTADOS

Os resultados oriundos das metas propostas para este trabalho são apresentados em 4 partes:

4.1 Artigo de divulgação científica: O Papel de Carboidratos na Função Proteica: O Caso das PGHSs

4.2 Manuscrito: The calcium goes meow: effects of ions and glycosylation on Fel d 1, the major cat allergen

4.3 Manuscrito: Conformacional characterization of fucosyltransferase 8 (FUT8) in solution

4.4 Caracterização da glicobiologia estrutura de C2GnT-L (resultados preliminares).

4.1 Artigo de divulgação científica: O Papel de Carboidratos na Função Proteica: o caso das PGHSs

O Papel de Carboidratos na Função Proteica: o caso das PGHSs

Sachett LG, Verli H

Revista Virtual de Química, 2011, 3(4), 275-285.



Artigo

O Papel de Carboidratos na Função Proteica: O Caso das PGHSs

Sachett, L. G.; Verli, H.*

Rev. Virtual Quim., 2011, 3 (4), 275-285. Data de publicação na Web: 25 de junho de 2011

<http://www.uff.br/rvq>

The Role of Carbohydrates on Protein Function: The Case of PGHSs

Abstract: It is estimated that about half of all existing proteins are glycosylated. When bound to proteins, carbohydrates are able to affect several properties of such molecules, including physicochemical and biological properties, and therefore they became essential for understanding its function and its therapeutical modulation. One of many strategies to obtain information on the role of a glycoprotein's saccharidic moiety involves the tridimensional study of such macromolecules. This can be achieved using molecular modeling, specifically, molecular dynamics simulations. Therefore, our research group has been dedicated to the development and validation of low computational cost approaches, based on free of charge tools, capable of contribute to the understanding of biological phenomena. These procedures are presented next, using Prostaglandin Endoperoxide Synthase 1 and 2 as example.

Keywords: glycoproteins; molecular dynamics; cyclooxygenase

Resumo

Estima-se que em torno de metade das proteínas existentes na natureza sejam glicosiladas. Ao ligarem-se às proteínas, os carboidratos podem alterar diversas de suas propriedades, tanto físico-químicas quanto biológicas, tornando-se fundamentais no entendimento de suas funções e, como consequência, na sua modulação com finalidades terapêuticas. Uma das formas de obtermos informações acerca do papel da porção sacarídica de uma glicoproteína envolve o estudo da estrutura tridimensional destas biomacromoléculas, por exemplo, através de cálculos de modelagem molecular, particularmente simulações de dinâmica molecular.

Nosso grupo de pesquisas vem se dedicando ao desenvolvimento e a validação de abordagens de baixo custo computacional e baseadas em ferramentas gratuitas capazes de contribuir para o entendimento de fenômenos biológicos. Estes procedimentos serão apresentados a seguir, tomando como exemplo as enzimas denominadas Prostaglandinas Endoperoxído Sintases, em suas isoformas 1 e 2.

Palavras-chave: glicoproteínas; dinâmica molecular; ciclo-oxigenases.



* Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Prédio 43431, sala 202, CP 15005, Porto Alegre 91500-970, RS, Brasil.

✉ hverli@cbiot.ufrgs.br

O Papel de Carboidratos na Função Proteica: O Caso das PGHSs

Liana G. Sachett, Hugo Verli*

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

* hverli@cbiot.ufrgs.br

Recebido em 21 de maio de 2011. Aceito para publicação em 14 de junho de 2011

1. Introdução

2. Como obter modelos atômicos para glicoproteínas?

2.1. Estrutura Proteica

2.2. Fração sacarídica

3. Papel da glicosilação na estrutura e função das PGHSs: impacto no planejamento de novos fármacos

4. Conclusões

1. Introdução

Carboidratos são moléculas de grande complexidade estrutural e conformacional. Suas unidades monoméricas podem se ligar umas às outras de diferentes maneiras, ao contrário de proteínas e ácidos nucleicos, que fazem apenas um tipo de ligação entre si. Outra diferença entre estes polímeros biológicos é que carboidratos podem ser altamente ramificados, enquanto proteínas e moléculas como DNA e RNA são quase exclusivamente lineares.¹

Os carboidratos são classificados de acordo com seu tamanho como monossacarídeos (uma única unidade de poli-hidroxi-aldeído ou cetona),

oligosacarídeos (cadeias curtas de unidades monossacarídicas) e polissacarídeos (polímeros de carboidratos de cadeias longas, geralmente com mais de 20 unidades monossacarídicas).² As ligações entre monossacarídeos, chamadas ligações glicosídicas, possuem duas possibilidades estereoquímicas, α ou β , (Figura 1) enquanto cada monossacarídeo pode adotar duas formas isoméricas possíveis, D ou L (para maiores informações acerca de nomenclatura e estereoquímica de carboidratos, consultar IUPAC³). Ainda, os resíduos sacarídicos podem apresentar diversas modificações covalentes, como acetilação, metilação e sulfatação, dentre outros.⁴

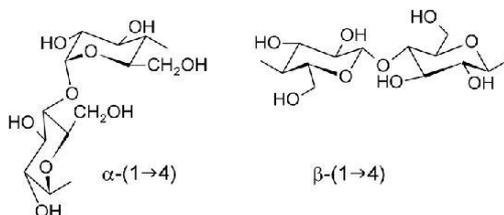


Figura 1. Exemplos de ligações glicosídicas do tipo α -(1 \rightarrow 4) e β -(1 \rightarrow 4)

Oligossacarídeos podem ser encontrados ligados a diversas classes de moléculas, como triterpenos, flavonoides, lipídeos e proteínas gerando, neste último caso, as denominadas glicoproteínas. Estas biomoléculas são encontradas majoritariamente no lado externo da membrana plasmática, na matriz extracelular e no sangue. No interior das células são encontradas em organelas específicas, como no complexo de Golgi, retículo endoplasmático, grânulos secretores e lisossomos.^{1,5} As glicoproteínas são fundamentais para muitos processos biológicos, incluindo resposta imune, crescimento celular, adesão entre células, degradação de coágulos sanguíneos e inflamação.^{1,6}

As estruturas sacarídicas podem se ligar às proteínas através do átomo de nitrogênio de resíduos de asparagina (N-glicosilação) e do átomo de oxigênio de resíduos de serina, treonina ou hidroxilisina (O-glicosilação). Adicionalmente, foram descritas C-, P- e S-glicosilações, embora não haja muitas informações sobre elas. A N-glicosilação apresenta um núcleo pentassacarídico característico e três tipos principais: oligomanose (apenas resíduos de manose adicionais ao núcleo), complexo (qualquer resíduo exceto manose adicionados ao núcleo) e híbrido (uma mistura dos dois tipos anteriores) (Figura 2).

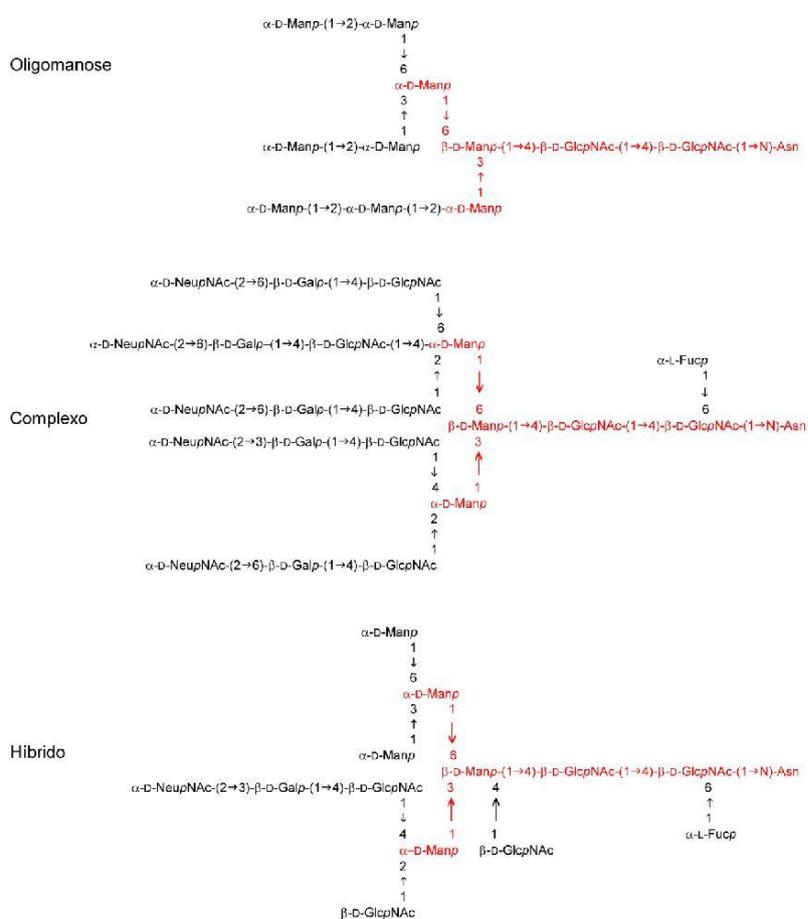


Figura 2. Exemplos dos três tipos principais de N-glicosilação. Em vermelho está o núcleo pentassacarídico comum a todas as N-glicanas

O tipo de glicosilação pode depender tanto da proteína envolvida quanto do tecido onde ocorre a expressão da glicoproteína. Contudo, ao contrário do que ocorre na síntese de cadeias polipeptídicas, a síntese de oligo e polissacarídeos gera seqüências de tamanhos variados, levando ao surgimento de populações de glicofórmos para uma mesma glicoproteína. Assim, uma glicoproteína deve ser visualizada como uma coleção de glicofórmos. Durante a adição das diferentes cadeias sacarídicas a uma determinada proteína, a glicosilação de um resíduo de aminoácido pode influenciar o processamento de outra glicosilação, principalmente através da limitação espacial ao acesso de glicosiltransferases.^{1,7}

Os oligossacarídeos ligados nas glicoproteínas são capazes de alterar diversas das propriedades proteicas, tanto físico-químicas quanto biológicas. Isso é particularmente importante considerando estimativas de que em torno de 50% das proteínas existentes na natureza sejam glicosiladas.⁸ Dentre as alterações físico-químicas podemos citar o aumento da resistência à proteólise e desnaturação, da viscosidade da solução, da solubilidade e a diminuição do ponto de congelamento da solução.⁹ Do ponto de vista biológico, a presença de carboidratos ligados a proteínas pode facilitar sua interação com chaperonas, modificando seu processo de enovelamento, alterar sua interação com outras proteínas, seu transporte e a formação de unidades multiméricas, além de ser capaz de alterar a atividade catalítica e o pH ótimo de enzimas.¹⁰

Um exemplo de glicoproteína de grande importância terapêutica envolve as enzimas Ciclo-oxigenases ou Prostaglandina Endoperóxido Sintases (PGHSs), enzimas homodiméricas que atuam sintetizando prostaglandinas a partir de ácido araquidônico. As prostaglandinas, por sua vez, atuam como moléculas sinalizadoras em processos como dor, febre, inchaço e inflamação. Existem duas isoformas bem caracterizadas destas enzimas: a PGHS-1, conhecida como isoforma constitutiva, e a PGHS-2, isoforma majoritariamente induzida. Ambas são N-glicosiladas e encontradas no lúmen do retículo endoplasmático e no envelope nuclear, apresentando atividades ciclo-oxigenase e peroxidase.¹¹ A partir da cristalografia de raios X obtida para ambas isoformas,^{12,13} pôde-se determinar as suas estruturas tridimensionais, sendo que cada monômero é composto por um domínio N-terminal do tipo fator de crescimento epidermal (EGF), um domínio de ligação à membrana (MBD, do inglês *membrane binding domain*) e um domínio catalítico C-terminal. Ambas PGHSs são N-glicosiladas nos resíduos de asparagina

Asn68, Asn144 e Asn410.¹⁴ As glicanas encontradas nestes sítios são do tipo oligomanose^{15,16} e fazem parte do processo de controle de qualidade para o correto enovelamento das enzimas.¹⁴

A estrutura dos oligossacarídeos presentes em glicoproteínas pode ser determinada experimentalmente por metodologias como a difração de raios-x e a ressonância magnética nuclear (RMN). Entretanto, se observarmos as estruturas de glicoproteínas depositadas no banco de dados *Protein Data Bank* (PDB, www.rcsb.org/pdb) obtidas por estes métodos, observamos que as estruturas sacarídicas geralmente estão incompletas. Isso se deve, principalmente, à alta flexibilidade das glicanas, que impede tanto a cristalização quanto a aquisição de densidade eletrônica adequada à solução das estruturas.¹⁷ Mesmo quando conseguimos ser obtidas por difração de raios-x, estas estruturas sacarídicas são de baixa qualidade, e estima-se que um terço dos dados para carboidratos encontrados no PDB contenha erros significativos de estereoquímica, nomenclatura e concordância com mapas de densidade eletrônica.¹⁸⁻²⁰ Ainda, frequentemente a porção sacarídica de glicoproteínas é removida para facilitar a cristalização de proteínas, de forma que muitas glicoproteínas tem sua estrutura tridimensional obtida na ausência de sua porção sacarídica²¹. Outro motivo para a ausência de carboidratos na estrutura de glicoproteínas se deve ao modo pelo qual estas moléculas são obtidas para cristalização, por exemplo, através de expressão recombinante em bactérias como *Escherichia coli*, conhecidas como incapazes de adicionarem carboidratos a resíduos de asparagina.¹⁸

Da mesma maneira, as estruturas obtidas por RMN também têm dificuldades em descrever corretamente carboidratos, já que elas representam uma média dos estados conformacionais existentes em solução, e a alta flexibilidade das glicanas produz um conjunto complexo de conformações, com a possibilidade de coexistirem múltiplos conformeros simultaneamente em solução.^{17,22-24} Como consequência, a média conformacional assim obtida pode não carregar consigo nenhuma ou pouca informação biologicamente relevante. A espectrometria de massas, por sua vez, é uma técnica capaz de descrever a composição dos oligossacarídeos contornando parte destas limitações,²⁵ embora não ofereça informações sobre a organização tridimensional dos compostos analisados.

A tridimensionalidade de estruturas sacarídicas é determinada, principalmente, pela geometria das ligações entre dois carboidratos vizinhos (ligação glicosídica). Esta, por sua vez, é principalmente

determinada pelos tipos de resíduos ligados, pelo tipo de ligação envolvida e pela interação destes resíduos com o solvente. Considerando estes aspectos, uma técnica adicional para estudo da estrutura tridimensional de carboidratos, ligados ou não a proteínas, envolve cálculos de modelagem molecular. Diversas abordagens vêm sendo empregadas na literatura para tal, tais como cálculos *ab initio*, cálculos de mecânica molecular, simulações por dinâmica molecular, *simulated annealing* e algoritmos genéticos, dentre outros.^{26,27}

Neste contexto, nosso grupo de pesquisas vem se dedicando ao desenvolvimento e validação de abordagens prontamente disponíveis à comunidade científica, de baixo custo computacional e baseadas em ferramentas gratuitas, capazes de contribuir no entendimento de fenômenos biológicos.²⁸ Estes procedimentos serão apresentados a seguir, tomando como exemplo as PGHSs em suas isoformas 1 e 2. As ferramentas empregadas incluem o campo de força GROMOS96,²⁹ o pacote de simulação GROMACS³⁰ e os servidores *Glycosciences*³¹ e PRODRG.³²

2. Como obter modelos atomísticos para glicoproteínas?

Em linhas gerais, é possível inferir-se a estrutura de uma glicoproteína completa, mesmo na ausência de informações tridimensionais prévias para a parte sacarídica, através do emprego de ferramentas computacionais.³³ Para tal, podemos seguir as seguintes etapas principais: 1) obtenção da estrutura tridimensional da parte proteica; 2) obtenção da estrutura tridimensional da parte sacarídica; 3) ligação das partes proteica e sacarídica e avaliação das potenciais interferências mútuas entre carboidratos e aminoácidos (para auxílio no entendimento das metodologias de modelagem molecular empregadas, ver Sant'Anna, 2009²⁶).

2.1. Estrutura proteica

Para obtenção da informação atomística da estrutura proteica, podemos empregar dados oriundos, principalmente, mas não exclusivamente, de: a) experimentos de difração de raios-X, b) experimentos de RMN, c) modelagem comparativa, d) modelagem por *folding recognition* e e) modelagem *ab initio*. Nos dois primeiros casos, os dados encontram-se disponíveis em repositórios públicos,

dentre os quais o principal é o PDB. Para os demais, programas e servidores gratuitos como Mholline para modelagem comparativa (<http://www.mholline.lncc.br/>), Robetta para modelagem *ab initio* (<http://robetta.bakerlab.org/>) e Phyre para *folding recognition* (<http://www.sbg.bio.ic.ac.uk/~phyre/>), dentre outros, estão disponíveis. No presente caso, as estruturas das enzimas foram obtidas através do banco de dados PDB, sob códigos 1Q4G e 1CVU para PGHS-1 e PGHS-2, respectivamente.

2.2. Fração sacarídica

A abordagem empregada para construção de estruturas oligossacarídicas envolve as seguintes etapas: a) construção de mapas de contorno relaxados para cada ligação glicosídica; b) refinamento de cada mínimo de energia através de simulações de dinâmica molecular; c) identificação dos estados conformacionais majoritários em solução a serem empregados na d) construção das porções sacarídicas. Esta estratégia visa fornecer modelos tridimensionais de carboidratos representativos de seus estados conformacionais em soluções biológicas. Adicionalmente, a mesma abordagem pode ser empregada em outros meios, como soluções não-biológicas para o estudo de reações químicas.

Este procedimento, embora independa de informações experimentais prévias sobre a estrutura tridimensional da parte sacarídica da glicoproteína, depende do conhecimento da sequência de resíduos de carboidratos ligados à proteína. Para as estruturas de glicosilação presentes na PGHS-1, dados de RMN indicam a existência apenas de resíduos de N-acetilglicosamina (D-GlcNAc) e manose (D-Manp) (portanto, uma glicana do tipo oligomanose), apresentando uma variação de 6 a 9 resíduos de manose em decorrência da existência de múltiplas glicofomas.¹⁵ Para as glicanas presentes na PGHS-2, dados de espectrometria de massas sugerem a existência de três glicofomas distintas, todas do tipo oligomanose, contendo de 6 a 10 resíduos de manose.¹⁶ Também existem dados para as glicanas das PGHSs no banco de dados PDB, embora todos estejam incompletos; a estrutura com maior número de resíduos de manose apresenta apenas três unidades deste carboidrato. A partir destas informações, tornou-se possível construir um modelo para os três sítios de glicosilação encontrados nas PGHSs, em concordância com os dados experimentais atualmente disponíveis para estas glicoproteínas (Figura 3A) (para detalhes adicionais ver Fernandes et

al., 2010³³).

Uma vez definida a estrutura sacarídica, torna-se necessário identificar sua conformação em um determinado meio, em geral aquoso quando estudamos moléculas em meio biológico. Uma das formas de obter tais informações é através dos denominados mapas de contorno relaxados. Estes mapas são análogos a mapas de Ramachandran³⁴ e descrevem mínimos locais de energia para cada ângulo de diedro analisado. Assim, para definição da

conformação adotada pela parte sacarídica das PGHSs cada dissacarídeo contido na estrutura apresentada na Figura 3A foi submetido ao cálculo de mapas de contorno relaxados, em torno da ligação glicosídica. Isso foi feito girando-se os ângulos de torção ϕ e ψ (Figura 4) de -180° até 150° , em passos de 30° , totalizando assim 144 confôrmeros para cada ligação entre dois monossacarídeos (Figura 3).

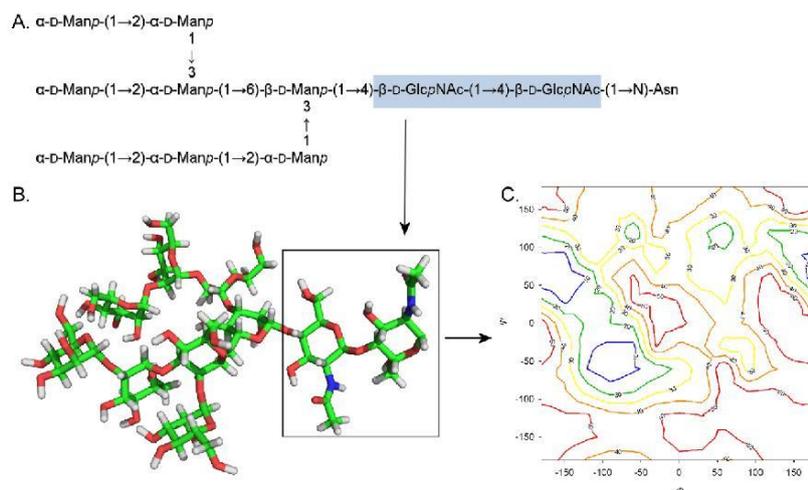


Figura 3. A. Estrutura do oligossacarídeo presente nos sítios de glicosilação das PGHSs. A ligação $\beta(1\rightarrow4)$ entre dois resíduos de D-GlcNAcp está destacada. B. Estrutura tridimensional do mesmo oligossacarídeo presente em A. C. Mapa de contorno obtido para a ligação glicosídica destacada nos itens anteriores, mostrando as conformações de menor energia para esta ligação (regiões em azul)

A orientação relativa de um par de resíduos de carboidratos é descrita, para diferentes tipos de ligação, por dois ou três ângulos de torção na ligação glicosídica (Figura 4). Para uma ligação $(1\rightarrow X)$, onde 'X' é 2, 3, 4 ou 6 para ligações $(1\rightarrow 2)$, $(1\rightarrow 3)$, $(1\rightarrow 4)$ e $(1\rightarrow 6)$ respectivamente, os ângulos ϕ e ψ são definidos como:

$$\phi = \text{O5-C1-OX-CX}$$

$$\psi = \text{C1-OX-CX-C(X-1)}$$

Para a ligação $(1\rightarrow 6)$ o ângulo ω é definido como:

$$\omega = \text{O6-C6-C5-C4}$$

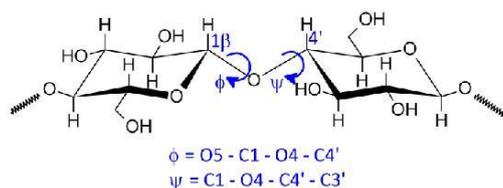


Figura 4. Ângulos de torção dos diedros ϕ e ψ para uma ligação do tipo $\beta(1\rightarrow 4)$ de um dissacarídeo

Um aspecto importante na obtenção de mapas de contorno é que os mesmos são usualmente obtidos no vácuo. As conformações de menor energia assim obtidas, portanto, podem apresentar menor correspondência com fenômenos que ocorram em meio biológico. Desta forma, buscando contornar esta limitação ao adicionar efeitos do solvente na geometria das moléculas em estudo, os diferentes estados de mínimo de energia obtidos dos mapas de contorno foram submetidos a simulações de dinâmica molecular, na presença de moléculas de água. Tais

simulações permitem-nos caracterizar a abundância relativa dos diferentes estados conformacionais adotados por cada ângulo de diedro em uma ligação glicosídica. Estas geometrias, apresentadas na Tabela 1, foram empregadas na construção das glicanas que, em seguida, podem ser ligadas a estrutura da proteína empregando o portal *Glycosciences* (<http://www.glycosciences.de/>)³¹ e, assim, possibilitar a obtenção de um modelo completo para a glicoproteína de interesse.

Tabela 1: Preferências conformacionais dos dissacarídeos presentes nas PGHSs quando simulados livres em solução

Dissacarídeo	Ângulo (°)		
	ϕ	ψ	ω
β -D-GlcNAc-(1→4)- β -D-GlcNAc	-80 ± 24	115 ± 16	-
β -D-Manp-(1→4)- β -D-GlcNAc	-160 ± 21	90 ± 18	-
α -D-Manp-(1→3)- α -D-Manp	160 ± 21	80 ± 21	-
α -D-Manp-(1→6)- β -D-Manp	80 ± 23	-160 ± 30	60 ± 17
α -D-Manp-(1→2)- α -D-Manp	80 ± 25	-135 ± 36	-

Dados expressos em média ± desvio padrão, para o estado conformacional majoritário de cada ângulo de diedro, obtidos a partir de dissacarídeos simulados livres em solução. Estas geometrias são empregadas como pontos de partida para estudos de DM de glicoproteínas. Para detalhes adicionais ver Pol-Fachin & Verli, 2011²⁸

3. Papel da glicosilação na estrutura e função das PGHSs: impacto no planejamento de novos fármacos

A partir da obtenção de um modelo para a estrutura completa de uma glicoproteína torna-se possível estudar o papel da parte sacarídica na estrutura, dinâmica e função proteicas. Para proteínas que são também receptores-alvo de fármacos, a obtenção de modelos mais precisos para a descrição de suas propriedades pode ter impacto direto no sucesso de esforços para obtenção de novos agentes terapêuticos. Uma das estratégias mais amplamente empregadas na literatura para obtenção destas informações, assim como para refinamento do modelo obtido, é a dinâmica molecular. Esta metodologia permite a caracterização da flexibilidade

e comportamento conformacional de moléculas e, particularmente no caso das glicoproteínas, o entendimento do papel da parte sacarídica na dinâmica de sua parte proteica.

Tal potencialidade das simulações de dinâmica molecular no entendimento da flexibilidade de compostos pode ser ilustrada na Figura 5, onde são sobrepostos mapas de contorno obtidos no vácuo aos estados conformacionais populados por diferentes ligações glicosídicas das glicanas ligadas às PGHSs. A maioria destes estados ocorre dentro ou nas proximidades de regiões de menor energia descritas nos respectivos mapas. Contudo, regiões adjacentes ou mesmo de maior energia são também populadas em solução, apontando para o papel do solvente na estabilização de diferentes estados conformacionais. Adicionalmente, podemos observar que o arcabouço glicoproteico é capaz de aumentar a rigidez das

ligações glicosídicas em relação à mesma ligação em dissacarídeos isolados (pontos vermelhos *versus* pontos azuis). Isto ocorre, principalmente, através do surgimento de interações intramoleculares entre os resíduos de carboidratos e de aminoácidos nas glicoproteínas. Contudo, estas interações tendem a

não promover grandes mudanças conformacionais nas ligações glicosídicas, isto é, não promovem novos estados conformacionais, mas predominantemente estabilizam estados já encontrados em solução para dissacarídeos isolados.³³

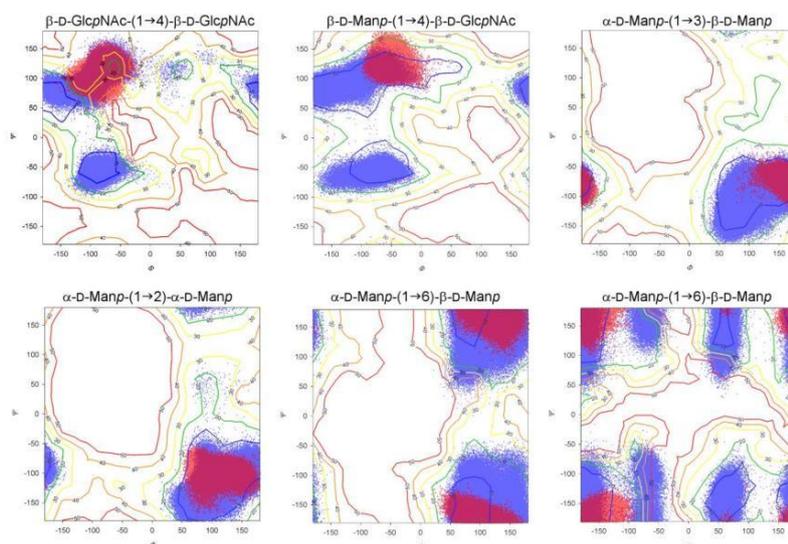


Figura 5. Mapas de contorno para cada uma das ligações glicosídicas contidas nas PGHSs. Cada mapa está sobrepostos às conformações populadas por cada ligação durante simulações de dinâmica molecular, quando na forma de dissacarídeos livres em solução (azul) e quando inseridos no ambiente glicoproteico (vermelho)

Uma das formas pelas quais carboidratos podem interferir na dinâmica proteica se dá através do enrijecimento da cadeia polipeptídica. Este tipo de alteração já havia sido descrito em trabalhos prévios no nosso grupo de pesquisa na proteína reguladora do complemento humana CD59, no domínio de adesão da proteína CD2, na subunidade a da gonadotrofina coriônica e também no domínio tipo EGF no fator VII de coagulação.²⁴ Nas PGHSs, também observamos um enrijecimento, na presença de glicosilação, no domínio EGF (Figura 6). Adicionalmente, dados obtidos por outros grupos têm demonstrado este tipo de efeito de enrijecimento do esqueleto peptídico na presença de glicanas, principalmente pelo aumento no número de interações polares entre soluto e solvente através dos resíduos de carboidratos adicionados, restringindo assim sua mobilidade.³⁵

Outra forma pela qual a estrutura de carboidratos pode interferir na estrutura proteica se dá através da

sua capacidade de cobrir uma ampla área da superfície proteica (Figura 7). Isso pode afetar a estrutura proteica de diversas maneiras, pois uma quanto maior a estrutura sacarídica, maior a sua capacidade de influenciar a conformação de diferentes regiões da glicoproteína, seja através de interações intramoleculares seja com o solvente. Isso também pode permitir que eles cubram áreas funcionalmente importantes, como sítios ativos, modulem interações de glicoconjugados com outras moléculas e afetem a taxa de ocorrência de mudanças conformacionais. Desta forma, carboidratos podem interferir diretamente, por exemplo, no acesso a sítios de ligação, na ativação proteica, na transmissão de efeitos conformacionais e na alteração na movimentação de alças. Adicionalmente, a glicosilação anormal de proteínas é diagnóstico de muitas doenças, incluindo artrite reumatóide e câncer.¹

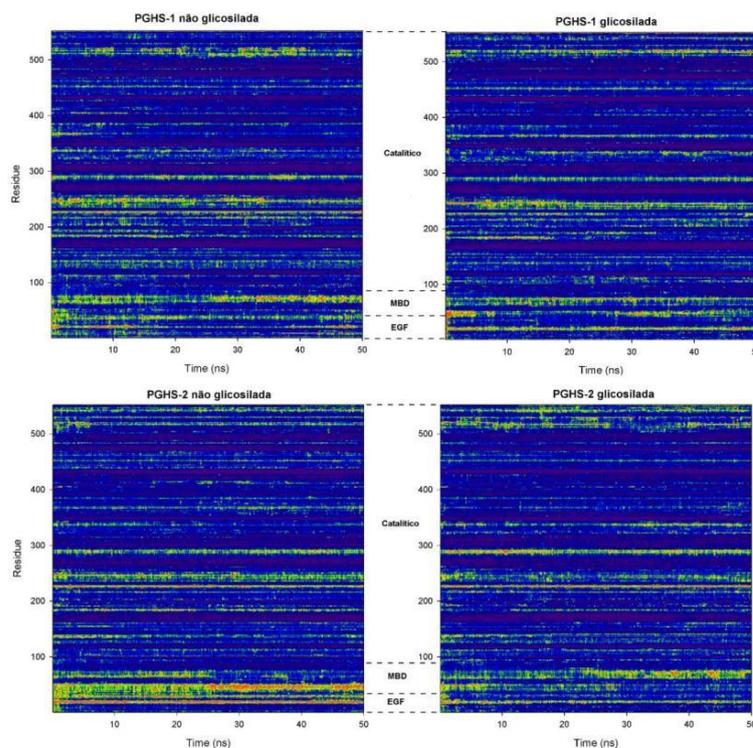


Figura 6. Dados de RMSF para as enzimas PGHS-1 e PGHS-2 glicosiladas e não glicosiladas. Este dado demonstra a variação na flexibilidade de cada resíduo durante o tempo de simulação. Regiões em laranja ilustram as maiores flexibilidades, enquanto regiões em azul pertencem a áreas mais rígidas

Tais influências da glicosilação nas propriedades proteicas podem impactar diretamente vários processos fisiológicos, particularmente aqueles dependentes de interações específicas proteína-proteína. De fato, interações entre proteínas e carboidratos (contidos ou não em glicoproteínas) são fundamentais para o processo de reconhecimento

entre células. Como consequência, o conhecimento da conformação de carboidratos torna-se fundamental para o entendimento da resposta biológica de interesse e sua posterior modulação terapêutica ao oferecer informações sobre os requisitos estruturais para o reconhecimento fármaco-receptor.

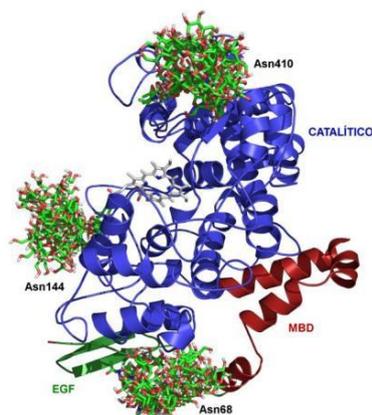


Figura 7. Sobreposição dos passos de uma simulação de dinâmica molecular (a cada 10 ns, num total de 50 ns de simulação) da enzima PGHS-2, contendo estruturas do tipo oligomanose nos três sítios de N-glicosilação, demonstrando assim a alta flexibilidade desta porção sacarídica

4. Conclusões

A aquisição de modelos confiáveis de estrutura tridimensional de carboidratos em sistemas biológicos complexos, como glicoproteínas, tem sido um desafio tanto para os métodos experimentais quanto para os teóricos. Enquanto a DM tem sido considerada progressivamente uma boa estratégia para obter descrições conformacionais de moléculas em solução, muito utilizada para descrever proteínas, ácidos nucleicos e membranas, seu uso para representar glicoproteínas é muito menos frequente. Isso se deve principalmente à falta de parametrização de carboidratos, tarefa a qual nosso grupo de pesquisa tem se dedicado.

A descrição conformacional de glicanas pertencentes a sistemas biológicos, em seus diferentes níveis de complexidade, utilizando pacotes de simulação moleculares rápidos e gratuitos, pode ser realizada, conforme descrito. Os mapas de contorno relaxados obtidos para os dissacarídeos resultaram em conformações de menor energia que representaram um ponto de partida adequado para os demais refinamentos por dinâmica molecular do comportamento conformacional dos dissacarídeos e ligações glicosídicas. A combinação de simulações no vácuo e em solução é capaz de oferecer uma base adequada para um conjunto conformacional de ligações glicosídicas, baseado apenas na sequência do sacarídeo e na ausência de informações experimentais prévias. Esta abordagem, portanto, é

uma estratégia potencial para a construção de modelos tridimensionais de glicoproteínas, como exemplificado para as enzimas PGHS-1 e PGHS-2, auxiliando no entendimento do papel biológico de glicoconjugados em nível atômico.

A análise de estruturas oligossacarídicas por dinâmica molecular se apresenta, assim, como um método apropriado para determinar as conformações possíveis de um oligossacarídeo, sendo capaz de facilitar o planejamento de análogos de carboidratos específicos, capazes de inibir interações carboidrato-proteína, e assim desenvolver novos agentes terapêuticos para tratamento de doenças humanas, como artrite reumatóide, doenças priônicas, câncer e desordens congênitas de glicosilação.

Agradecimentos

Conselho Nacional de Desenvolvimento Científico e Tecnológico, MCT e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brasil.

Referências Bibliográficas

- ¹ Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683. [[CrossRef](#)] [[PubMed](#)]

Sachett, L. G.; Verli, H.



- ² Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E.; *Essentials of Glycobiology*, 2a ed., Cold Spring Harbor: New York, 2009.
- ³ I.U.P.A.C.-I.U.B. *Commission on Biochemical Nomenclature Pure Applied Chemistry* **1996**, 68, 1919.
- ⁴ Pilobello, K. T.; Mahal, L. K. *Curr. Opin. Chem. Biol.* **2007**, 11, 300. [[CrossRef](#)] [[PubMed](#)]
- ⁵ Gahmberg, C. G.; Tolvanen, M. *Trends Biochem. Sci.* **1996**, 8, 308. [[PubMed](#)]
- ⁶ Ohtsuba, K.; Marth, J. D. *Cell* **2006**, 126, 855. [[CrossRef](#)] [[PubMed](#)]
- ⁷ Wormald, M. R.; Wooten, E. W.; Bazzo, R.; Edge, C. J.; Feinstein, A.; Rademacher, T. W.; Dwek, R. A. *Eur. J. Biochem.* **1991**, 198, 131. [[CrossRef](#)] [[PubMed](#)]
- ⁸ Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, 1473, 4. [[PubMed](#)]
- ⁹ Varki, A. *Glycobiology* **1993**, 3, 97. [[CrossRef](#)] [[PubMed](#)]
- ¹⁰ Sears, P.; Wong, C. -H. *Cell. Mol. Life Sci.* **1998**, 54, 223. [[CrossRef](#)]
- ¹¹ DeWitt, D. L.; Rollins, T. E.; Day, J. S.; Gauger, J. A.; Smith, W. L. *J. Biol. Chem.* **1881**, 256, 10375. [[PubMed](#)]
- ¹² Picot, D.; Loll, P. J.; Garavito, R. M. *Nature* **1994**, 367, 243. [[CrossRef](#)] [[PubMed](#)]
- ¹³ Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; Isakson, P. C.; Stallings, W. C. *Nature* **1996**, 384, 644. [[CrossRef](#)] [[PubMed](#)]
- ¹⁴ Otto, J. C.; DeWitt, D. L.; Smith, W. L. *J. Biol. Chem.* **1993**, 268, 18234. [[PubMed](#)]
- ¹⁵ Mutsaers, J. H. G. M.; van Halbeek, H.; Kamerling, J. P.; Vliegenthart, J. F. G. *Eur. J. Biochem.* **1985**, 147, 569. [[CrossRef](#)]
- ¹⁶ Nemeth, J. F.; Hochesang, G. P.; Marnett, L. J.; Caprioli, R. M. *Biochemistry* **2001**, 40, 3109. [[CrossRef](#)] [[PubMed](#)]
- ¹⁷ Woods, R. J. *Glycoconjugate J.* **1998**, 15, 209. [[CrossRef](#)] [[PubMed](#)]
- ¹⁸ Lütteke, T.; Frank, M.; von der Lieth, C. -W. *Carbohydr. Res.* **2004**, 339, 1015. [[CrossRef](#)] [[PubMed](#)]
- ¹⁹ Crispin, M.; Stuart, D. I.; Jones, E. Y. *Nat. Struct. Mol. Biol.* **2007**, 14, 354. [[CrossRef](#)]
- ²⁰ Berman, H. M.; Henric, K.; Nakamura, H.; Markley, J. *Nat. Struct. Mol. Biol.* **2007**, 14, 354. [[CrossRef](#)]
- ²¹ McPherson, A. *Crystallization of Biological Macromolecules*, Cold Spring Harbor: New York, **1999**.
- ²² Cumming, D. A.; Carver, J. P. *Biochemistry* **1987**, 26, 6664. [[CrossRef](#)] [[PubMed](#)]
- ²³ Wormald, M.; Petrescu, A. -J.; Pao, Y. -L.; Glythero, A.; Elliot, T.; Dwek, R. A. *Chem. Rev.* **2002**, 102, 371. [[CrossRef](#)] [[PubMed](#)]
- ²⁴ Pol-Fachin, L.; Fernandes, C. L.; Verli, H. *Carbohydr. Res.* **2009**, 344, 491. [[CrossRef](#)] [[PubMed](#)]
- ²⁵ Pérez, S.; Mulloy, B. *Curr. Opin. Struct. Biol.* **2005**, 15, 517. [[CrossRef](#)] [[PubMed](#)]
- ²⁶ Sant'Anna, Carlos M. R. *Rev. Virtual Quim.* **2009**, 1, 49. [[Link](#)]
- ²⁷ Verli, H.; Guimarães, J. A. *Strategies for the determination of carbohydrates structure and conformation.*, 1a. ed., Kerala: Transworld Research Network, **2010**.
- ²⁸ Pol-Fachin, L.; Verli, H. *Mini-Rev. Org. Chem.*, **2011**. (No prelo)
- ²⁹ van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; *Simulation, Biomolecular. The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag: AG Zurich, Switzerland, 1996.
- ³⁰ van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, 26, 1701. [[CrossRef](#)] [[PubMed](#)]
- ³¹ Lütteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C. -W. *Glycobiology* **2006**, 16, 71R. [[CrossRef](#)] [[PubMed](#)]
- ³² Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, 60, 1355. [[CrossRef](#)] [[PubMed](#)]
- ³³ Fernandes, C. L.; Sachett, L. G.; Pol-Fachin, L.; Verli, H. *Carbohydr. Res.* **2010**, 345 663. [[CrossRef](#)] [[PubMed](#)]
- ³⁴ Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, 7, 95. [[CrossRef](#)] [[PubMed](#)]
- ³⁵ Perera, L.; Darden, T. A.; Pedersen, L. G. *J. Comput. Chem.* **2002**, 23, 35. [[CrossRef](#)] [[PubMed](#)]

4.2 Artigo: The calcium goes meow: effects of ions and glycosylation on Fel d 1, the major cat allergen

Rodrigo Ligabue-Braun, Liana Guimarães Sachett, Laércio Pol-Fachin, Hugo Verli

Submetido à revista Plos One em 1 de janeiro de 2015. Aceito em 11 de junho de 2015. Publicado em 2 de julho de 2015.

RESEARCH ARTICLE

The Calcium Goes Meow: Effects of Ions and Glycosylation on Fel d 1, the Major Cat Allergen

Rodrigo Ligabue-Braun¹*, Liana Guimarães Sachett¹*, Laércio Pol-Fachin², Hugo Verli¹*

1 Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil, **2** Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife, PE, Brazil

* These authors contributed equally to this work.

* hverli@cbiot.ufrgs.br



CrossMark
click for updates

OPEN ACCESS

Citation: Ligabue-Braun R, Sachett LG, Pol-Fachin L, Verli H (2015) The Calcium Goes Meow: Effects of Ions and Glycosylation on Fel d 1, the Major Cat Allergen. PLoS ONE 10(7): e0132311. doi:10.1371/journal.pone.0132311

Editor: Yaakov Koby Levy, Weizmann Institute of Science, ISRAEL

Received: January 1, 2015

Accepted: June 11, 2015

Published: July 2, 2015

Copyright: © 2015 Ligabue-Braun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Large simulation trajectories are available upon request.

Funding: This work was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul), and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, grant number APQ-0398-1.06/13).

Abstract

The major cat allergen, Fel d 1, is a structurally complex protein with two N-glycosylation sites that may be filled by different glycoforms. In addition, the protein contains three putative Ca^{2+} binding sites. Since the impact of these Fel d 1 structure modifications on the protein dynamics, physiology and pathology are not well established, the present work employed computational biology techniques to tackle these issues. While conformational effects brought upon by glycosylation were identified, potentially involved in cavity volume regulation, our results indicate that only the central Ca^{2+} ion remains coordinated to Fel d 1 in biological solutions, impairing its proposed role in modulating phospholipase A_2 activity. As these results increase our understanding of Fel d 1 structural biology, they may offer new support for understanding its physiological role and impact into cat-promoted allergy.

Introduction

Allergic diseases, especially those involving responses mediated by immunoglobulin E (IgE), are increasing in prevalence and becoming major public health issues [1–3]. This increase has been associated with a Westernized life style and urbanization, suggesting that pets could contribute to this scenario as sources of indoor allergens [1, 4]. The most implicated pets in allergic diseases are cats (*Felis domesticus*), which are present in up to 50% of homes [5]. The sensitization prevalence in adults is in the range of 10–15%, with symptoms varying from rhinoconjunctivitis to potentially life-threatening asthmatic exacerbations [6].

Among the few components of cat dander that can elicit IgE response, the Fel d 1 protein is considered the most potent [6]. Fel d 1, the major cat allergen, is a dimer of all-helical heterodimers, included in the secretoglobin family [7, 8]. Each dimer in the native heterotetramer is N-glycosylated [9], and crystallographic analyses of recombinant Fel d 1 revealed that each dimer has a cavity, possibly involved in the transport of an unknown molecule [7, 8]. For crystallization, a fused version of the protein was produced, in which chains 1 and 2 are linked, making the original tetramer (or dimer of dimers), a simple dimer of the fused protein [8] (Fig 1).

Competing Interests: The authors have declared that no competing interests exist.

The crystallographic study of this recombinant Fel d 1 tetramer also revealed three putative calcium-binding sites, one at the tetramerization interface and one on the lateral of each dimer [8]. The lateral binding sites involve O atoms from Asp46 and Met49 (main chain) and three water molecules within coordination distance from the Ca^{2+} , while the binding site at the interface comprises OD1 atoms from Asn89 (from both chains A and B) and Asp130 (chain B), a carbonyl from Ile125 (chain A) and three water molecules [8] (Fig 1). The non-interfacial binding sites are proposed to act on allergic responses by modulating phospholipase A_2 activity via calcium sequestration, a property related to uteroglobins [8, 10]. The interaction of Fel d 1 and calcium is however subject of debate, since for uteroglobin (a member of the secretoglobulin protein family) some authors identified calcium binding, while others identified the opposite [10–13].

The importance of Fel d 1 glycosylation in protein structure and immune response has also been somewhat controversial. While some studies show that different glycosylation patterns do not affect IgE production *in vitro* [9, 14, 15], a most recent study demonstrated that the mannose receptor has an essential role in internalizing Fel d 1. Mannose receptor cysteine rich domain recognizes the carbohydrates in Fel d 1 and *in vivo* assays showed that knockout mice for mannose receptor produced lower levels of immunoglobulins E and G [16].

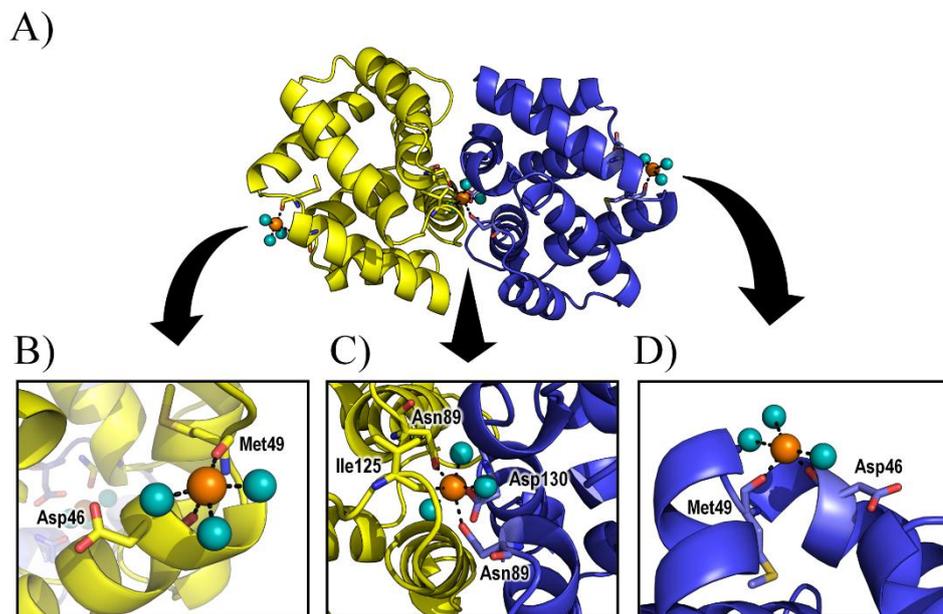
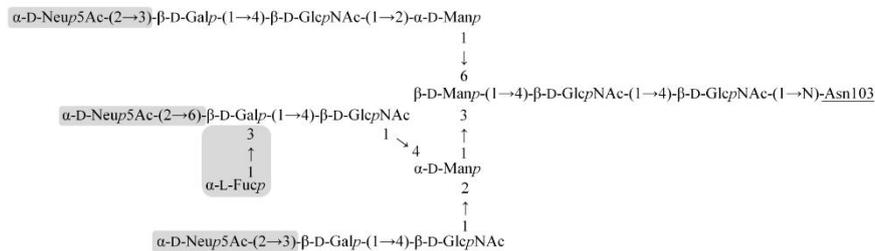


Fig 1. Fel d 1 Ca^{2+} binding sites. (A) Fel d 1 crystallographic structure highlighting the location of calcium ions; (B) Lateral Ca^{2+} binding site A; (C) Interfacial Ca^{2+} binding site B; (D) Lateral Ca^{2+} binding site B. Proposed coordination interactions based on the crystallographic structure are shown as dashed lines. Calcium ions are shown as orange spheres, water molecules are shown as cyan spheres, Chain A is shown in yellow, Chain B is shown in blue. (the orientation of the boxes is different from the one in (A) for clarity).

doi:10.1371/journal.pone.0132311.g001

A)



B)

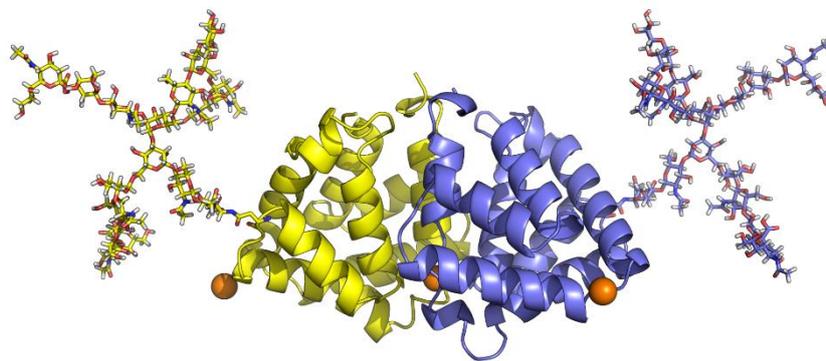


Fig 2. Fel d 1 glycosylation structures. (A) Full (largest) glycosylation structure of Fel d 1, shaded boxes indicate saccharides that are absent in the minimal (smallest) glycosylation structure. (B) Fel d 1 structure with full glycan chains attached. Calcium ions are shown as orange spheres, Chain A is shown in yellow, Chain B is shown in blue.

doi:10.1371/journal.pone.0132311.g002

The glycosylation pattern of Fel d 1 has been determined by mass spectrometry, revealing a series of possible glycoforms bound to the protein [14]. As observed for other glycosylated proteins, there is a minimal structure to which different oligosaccharides may be added, leading to variation in the glycan moiety, which is limited to the largest glycosylation tree identified. For the present work we employed the extremes of the glycosylation structure, analyzing the largest (or "full glycosylation") and the smallest ("minimal glycosylation") structures found by mass spectrometry [14] (Fig 2). Intermediate glycostructures have different numbers of galactose, mannose, and sialic acid residues.

Despite the extensive studies on Fel d 1-induced allergy in humans (as reviewed by [6, 17]), very little information is available regarding its interactions with calcium ions or the effect of glycosylation. Considering that these properties may be important not only for allergy studies but also for understanding the physiological role of this protein in felines, the present work employed computational biology techniques to tackle these issues. Effects brought upon by

glycosylation were identified, while the lateral calcium binding sites proposed by crystallography were not confirmed.

Results

Outline

In this work we performed molecular dynamics (MD) simulations of Fel d 1 using different force fields to evaluate a putative role of parameter differences in describing Ca^{2+} behavior in the protein environment. These molecular mechanics studies were complemented with semi-empirical calculations of the Ca^{2+} interaction with Fel d 1, providing electronic information regarding these ions. In a separate analysis, we inspected the role of two different glycosylation structures on the protein structure and dynamics. These glycoforms are the largest (full glycosylation) and smallest (minimal glycosylation) found on Fel d 1. Protein cavities and ligand-binding possibilities were also studied. The simulated systems and calculations are schematically listed below. Please refer to the Material and Methods section at the end of the article for version information, simulation details, and further references.

By MD we simulated the Fel d 1 dimer of fused dimers under three force fields (AMBER, CHARMM, GROMOS) for 200 ns, in order to guarantee that the Ca^{2+} orientation within Fel d 1 was not due to an specific set of parameters, as well as a simplified structural mockup of the crystallographic environment of Fel d 1 for 25 ns (GROMOS force field) in order to trace potential crystallographic contacts able to lock the Ca^{2+} in position. The role of glycan chains in Fel d 1 dynamics was evaluated through 200 ns simulations using GROMOS force field for two glycosylated systems (Fel d 1 in the largest and smallest glycoforms). After evaluation on the putative Ca^{2+} position dependence on force field parameters, the simulation of the glycoproteins employed GROMOS due to its robust set of parameters for glycoprotein simulations [18, 19], as well as to the lower computational cost associated to united atom force fields in comparison to all atom parameters. All MD simulations were carried out with the GROMACS simulation package. To include direct information on the ions degree of coordination, semi-empirical calculations of the Ca^{2+} binding sites were performed with MOPAC under three scenarios: with implicit solvent, without implicit solvent, and with the inclusion of crystallographic waters.

Implementation and validation of GROMOS 53A6glyc sialic acid parameter

In order to enable Fel d 1 glycoforms to be studied by MD simulations, and considering that parameters for neuraminic (sialic) acid residues were absent in the original GROMOS 53A6glyc implementation, we have constructed a topology for such monosaccharide in the current work (S1 Table). It was built based on ring and hydroxyl parameters for aldohexopyranoses [18] under GROMOS 53A6glyc, as well as N-acetyl and carboxylate parameters previously employed in GROMOS 45A4/53A6glyc force fields [19, 20]. Before employing such parameters for fully glycosylated Fel d 1 MD simulations, we have validated such topology by evaluating isolated sialic acid conformational ring pucker properties, as performed for other monosaccharides [18]. Thus, while metadynamics calculations indicated ${}^2\text{C}_5$ as the preferential ring pucker conformation for sialic acid, unbiased 1000 ns MD simulations confirmed such behavior, in accordance with previous observations (S1 Fig).

Table 1. Distances and bond orders measured by semi-empirical calculation.

Lateral Ca ²⁺ (A)		With implicit solvent		Without implicit solvent		With crystallographic waters	
	O atom	Distance (Å)	Bond order	Distance (Å)	Bond order	Distance (Å)	Bond order
Asp46A	Side chain	-	-	2.2	0.274	-	-
	Main chain	2.2	0.234	-	-	-	-
Met49A	Main chain	2.3	0.291	2.4	0.102 [†]	2.4	0.094 [†]
Thr50A	Side chain	2.3	0.230	-	-	-	-
	Side chain	-	-	2.3	0.313	2.3	0.279
Glu51A	Side chain	-	-	-	-	2.3	0.243
	Main chain	-	-	2.3	0.268	-	-
H ₂ O	Water 1	NA	NA	NA	NA	2.4	0.222
	Water 2	NA	NA	NA	NA	2.4	0.207
	Water 3	NA	NA	NA	NA	2.4	0.179 [†]
Central Ca ²⁺		With implicit solvent		Without implicit solvent		With crystallographic waters	
	O atom	Distance (Å)	Bond order	Distance (Å)	Bond order	Distance (Å)	Bond order
Asn89A	Side chain	2.2	0.255	2.3	0.238	2.3	0.208
Ile125A	Main chain	2.3	0.187 [†]	2.5 [‡]	0.086 [†]	-	-
Asn89B	Side chain	2.3	0.195 [†]	2.3	0.219	2.3	0.227
Ile125B	Main chain	-	-	2.3	0.182 [†]	-	-
	Side chain	2.3	0.248	2.3	0.268	2.3	0.278
Asp130B	Side chain	2.3	0.257	2.3	0.268	2.3	0.227
	Water 7	NA	NA	NA	NA	2.4	0.197 [†]
H ₂ O	Water 8	NA	NA	NA	NA	2.5 [‡]	0.122 [†]
	Water 8	NA	NA	NA	NA	2.5 [‡]	0.122 [†]
Lateral Ca ²⁺ (B)		With implicit solvent		Without implicit solvent		With crystallographic waters	
	O atom	Distance (Å)	Bond order	Distance (Å)	Bond order	Distance (Å)	Bond order
Asp46B	Side chain	-	-	2.3	0.192 [†]	2.2	0.226
	Main chain	2.3	0.223	2.3	0.238	-	-
Ala47B	Main chain	2.3	0.211	-	-	-	-
Met49B	Main chain	2.3	0.264	2.4	0.148 [†]	2.3	0.157 [†]
Thr50B	Main chain	-	-	2.5 [‡]	0.102 [†]	-	-
	Side chain	2.3	0.250	2.3	0.237	-	-
Glu51B	Side chain	-	-	2.3	0.296	-	-
	Water 4	NA	NA	NA	NA	2.3	0.250
H ₂ O	Water 5	NA	NA	NA	NA	2.3	0.301
	Water 6	NA	NA	NA	NA	2.3	0.261

[†]Bond orders of less than 0.2 are indicative of "no bond" [21]

[‡]Distances greater than reference distances for calcium coordination in proteins (2.35 Å-2.45 Å) [22].

doi:10.1371/journal.pone.0132311.t001

Analyzes of putative calcium binding sites

The three calcium binding sites proposed for Fel d 1 (Fig 1) were analyzed by semi-empirical calculations and molecular dynamics (MD) simulations. Three conditions were calculated semi empirically: with implicit solvent, without implicit solvent, and including crystallographic waters. The distance and bond orders obtained from these calculations (Table 1) indicate that the central Ca²⁺ is stable even if a little dislocated in its position, while the lateral ions are not bound and move from their proposed locations. MD simulations carried out in three force fields (AMBER, CHARMM, GROMOS) indicate the same dislodgment of the lateral Ca²⁺ ions (Table 2 and S2 Fig). To inspect if crystal packing effects would stabilize these ions, the

Table 2. Average distances measured for molecular dynamics simulations under different force fields (Å).

	*	AMBER	CHARMM	GROMOS	Crystal	Expected*
Lateral Ca ²⁺ (A)	Asp46A	37.0±1.0	11.1±1.5	33.7±13.3	13.5±14.8	2.4
	Met49A	36.8±8.9	10.0±1.3	37.3±13.7	14.2±15.4	2.3
	Asn89A	2.7±0.1	2.3±0.1	2.5±0.3	2.4±0.1	2.4
Central Ca ²⁺	Ile125A	2.7±0.1	4.3±0.7	7.2±0.8	4.2±1.0	2.2
	Asn89B	2.7±0.3	5.3±1.0	4.9±0.6	4.2±0.8	2.7
	Asp130B	2.6±0.1	2.3±0.3	2.7±0.2	5.2±0.2	2.2
Lateral Ca ²⁺ (B)	Asp46B	4.1±0.9	16.1±1.8	27.3±11.2	7.6±4.5	2.4
	Met49B	5.3±0.6	13.2±2.1	27.0±11.0	6.7±3.7	2.3

*residues making contact with calcium ions and their expected distances were taken from PDB ID 2EJN [8].

doi:10.1371/journal.pone.0132311.t002

neighboring asymmetric units from the crystallographic Fel d 1 were regenerated and simulated (Fig 3). There was no stabilization detected in this condition as well (Table 2).

Calcium effects on protein conformation

Considering that the stable binding of the central Ca²⁺ could be related to stabilization of the protein dimer, the interaction energy of the complex was also measured for the performed MD simulations (Table 3 and S3 Fig). There was a slight decrease in interaction energy (more favored interaction) in the calcium-free conditions when compared to calcium-bound conditions.

The MD simulated systems were generally stable (Fig 4) and had no difference regarding their stability with or without Ca²⁺ for simulations under AMBER and CHARMM force fields. In the simulations performed under GROMOS force field, however, there was a slight difference between metal-bound and metal free conditions combined with noticeable difference in stability in comparison with the other force fields. (Fig 4). Regarding local flexibility, we observed little differences between calcium-free and calcium-bound system (Fig 5).

Glycosylation effects

The minimally and fully glycosylated systems behaved differently. While the fully glycosylated protein had a conformational behavior equivalent to the unglycosylated forms in the same force field (GROMOS), the protein with minimal glycosylation had a greater divergence in terms of conformation (Fig 6).

Moreover, glycosylation seemed to increase flexibility around residue 75, in a loop region of chain A (but not on the equivalent region of chain B) (Fig 6). The minimal saccharide structure also caused an increase in flexibility in the N-terminus region of both chains. The secondary structure content was similar in all simulations, with a noted decrease in helical content for the systems with no Ca²⁺ and with the minimally glycosylated protein (S4 Fig). This decrease was associated with partial unfolding of helices I and II (first and second helices from the N-terminus), reflected in the increased flexibility observed in RMSF analyses (Fig 6). The analysis of secondary structure along the simulated time also indicates structural stability for (at least) the last 20 ns of all simulation with the GROMOS force field. It is interesting to note that the full glycosylation seemed to stabilize the Chain B structure, as revealed by RMSD analyses of each of the protein segments (half-monomers in the fused protein) (S5 Fig). This protein-segment analysis also confirmed the early stabilization of all chains in the Ca²⁺ system.

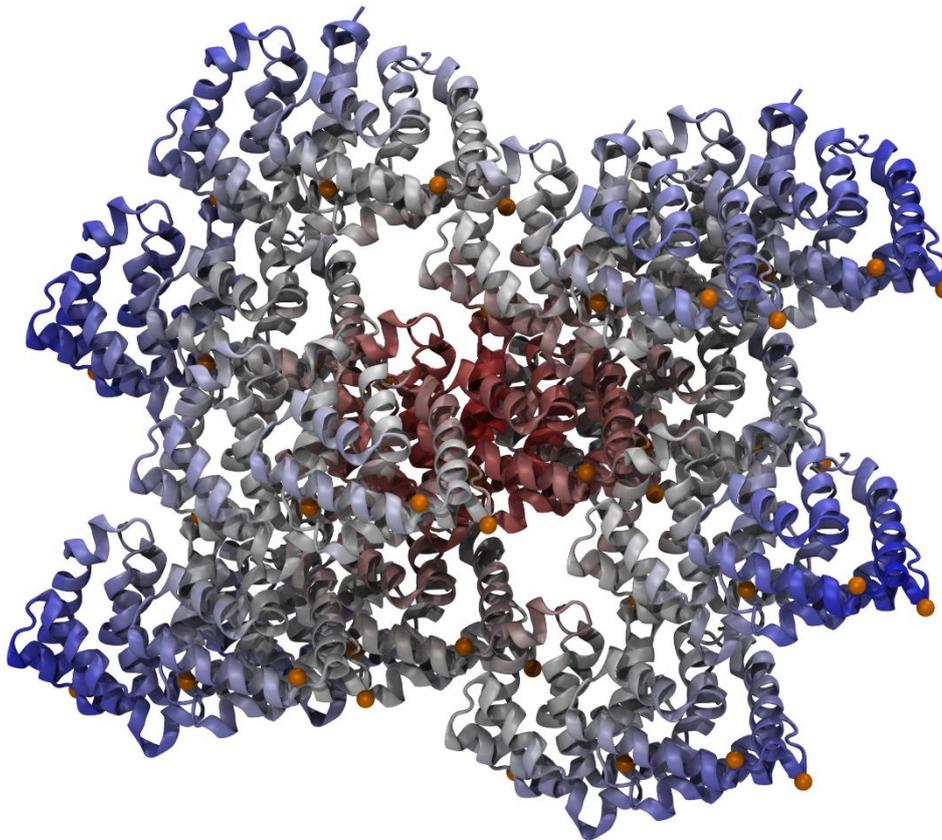


Fig 3. Fel d 1 in its crystallographic environment. Red depicts the central protein, while the white-to-blue gradient indicates surrounding protein molecules. Calcium ions are shown as orange spheres.

doi:10.1371/journal.pone.0132311.g003

Table 3. Interaction energies measured between Fel d 1 monomers during molecular dynamics simulations under different force fields (kJ/mol).

	AMBER	CHARMM	GROMOS
With Ca ²⁺ ions	-718.07±65.49	-599.44±69.71	-570.57±67.84
Without Ca ²⁺ ions	-728.87±69.52	-772.41±67.47	-617.71±87.07

doi:10.1371/journal.pone.0132311.t003

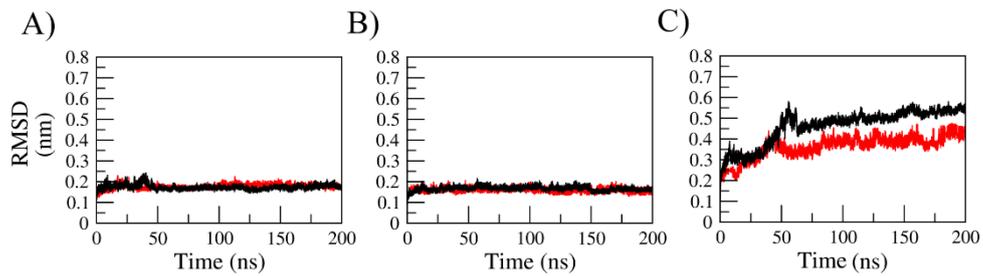


Fig 4. All-atom root mean square deviation (RMSD) measured for different force field simulations of Fel d 1. (A) AMBER, (B) CHARMM, (C) GROMOS. Color coding: Simulations without Ca^{2+} ions (black), simulations with Ca^{2+} ions (red).

doi:10.1371/journal.pone.0132311.g004

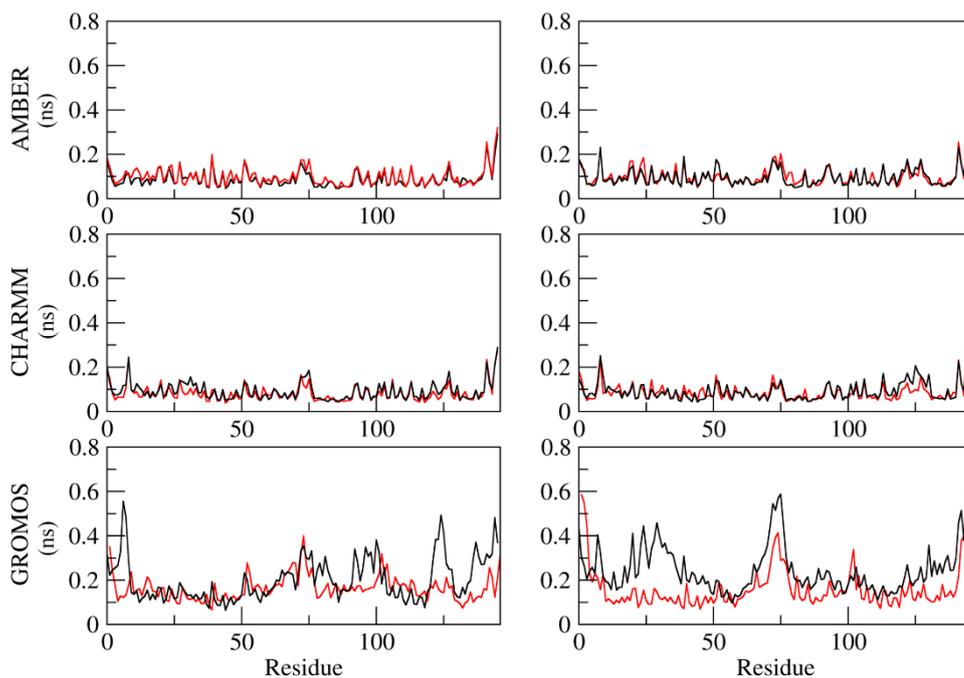


Fig 5. Root mean square fluctuations (RMSF) measured for different force field simulations of Fel d 1. Color coding: Simulations without Ca^{2+} ions (black), simulations with Ca^{2+} ions (red).

doi:10.1371/journal.pone.0132311.g005

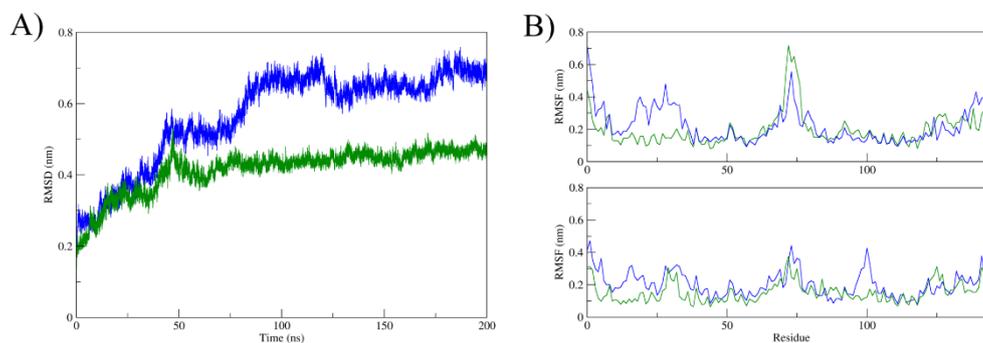


Fig 6. Structural flexibility of glycosylated Fel d 1. (A) All-atom root mean square deviation (RMSD) and (B) Root mean square fluctuations (RMSF) measured for the simulations with minimal (blue) and full glycosylations (green).

doi:10.1371/journal.pone.0132311.g006

Cavity assessment and ligand dockings

The two cavities detected in Fel d 1, which had different volumes in the crystallographic structure [8] were measured for all systems simulated with GROMOS force field. Their volume seems to oscillate periodically along the simulated time (S6 Fig). This is especially clear by analyzing the three main structural clusters found in each simulation (Table 4). The cavities detected in these clusters confirm clearly different cavity sizes for each Fel d 1 chain.

We studied Fel d 1 cavity binding capacity using molecular docking, revealing that cavities of both Fel d 1 chains (A and B) are able to interact with the ligands PCB (an steroid analog), progesterone and testosterone (Fig 7). In chain A, residues Val10, Leu61 and Phe84 presented hydrophobic interactions with all three ligands, and Asp130 was responsible for one hydrogen bond. In chain B, for all ligands, residues Phe13 and Phe84 were the ones making the hydrophobic contacts and Tyr21 was the residue responsible for one hydrogen bond.

Dynamical network and motion analyzes

Inspecting internal motion correlations for the simulated Fel d 1 systems with GROMOS force field, we were able to detect two major trends. First, there is conservation of residues acting as

Table 4. Cavity volume measured for the three main structural clusters from MD simulations under different conditions.

		Volume (Å ³)		
		Cluster 1	Cluster 2	Cluster 3
With Ca ²⁺	Chain A	4357	4133	4174
	Chain B	5145	4004	3783
Without Ca ²⁺	Chain A	4397	4590	3208
	Chain B	4035	4383	4454
Glycosylated (minimal)	Chain A	3133	3537	3634
	Chain B	2907	4367	4217
Glycosylated (full)	Chain A	3462	4032	3344
	Chain B	4607	4031	3549

doi:10.1371/journal.pone.0132311.t004

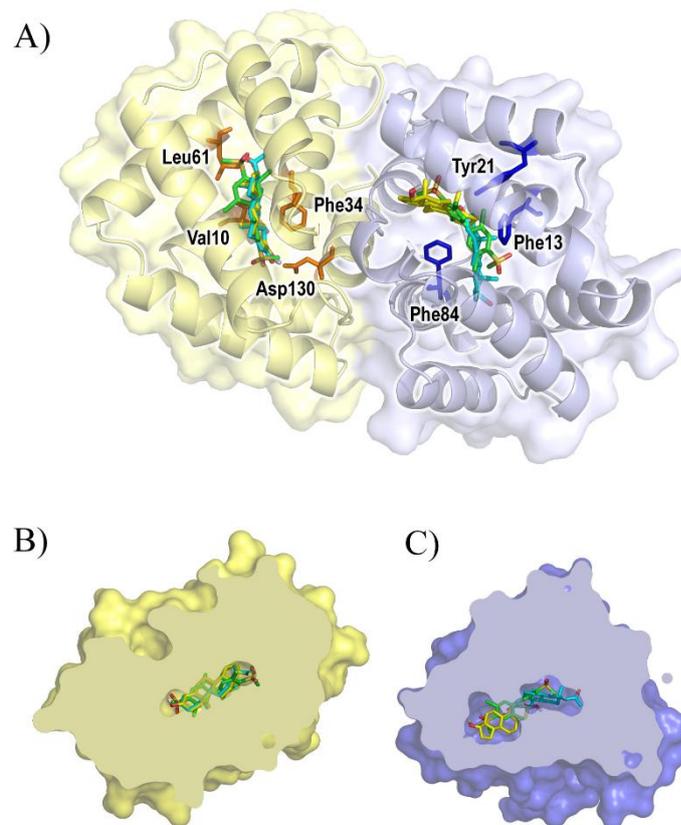


Fig 7. Docking of different ligands to Fel d 1. (A) Docking results overview, highlighting residues involved in binding interactions; Cut-through views of (B) chain A and (C) chain B, highlighting differences in cavity sizes. Cut-through views are rotated in respect to (A) for clarity; chain A shown in yellow and chain B shown in blue, testosterone shown in yellow, progesterone shown in cyan, PCB shown in green.

doi:10.1371/journal.pone.0132311.g007

critical nodes for motion transmission between chains A and B at the interaction interface, regardless of the presence of calcium or glycosylation (S2 Table). Second, there seems to be a hinge-like motion transfer between each original chain (i.e. each half of Fel d 1 crystallographic monomer) in the simulated systems (Fig 8). The observed critical nodes, also conserved, seem to act as unconventional hinges, and an oscillatory motion between these ‘flaps’ can be observed, especially for Chain A (S7 Fig), although the characteristic disulfide bridges of Fel d 1 may restrict this oscillation.

The motions of each protein chain in respect to each other and within each chain (Fig 9) reveal no clear pattern. The calcium-free system has diffuse motions, whereas the other systems

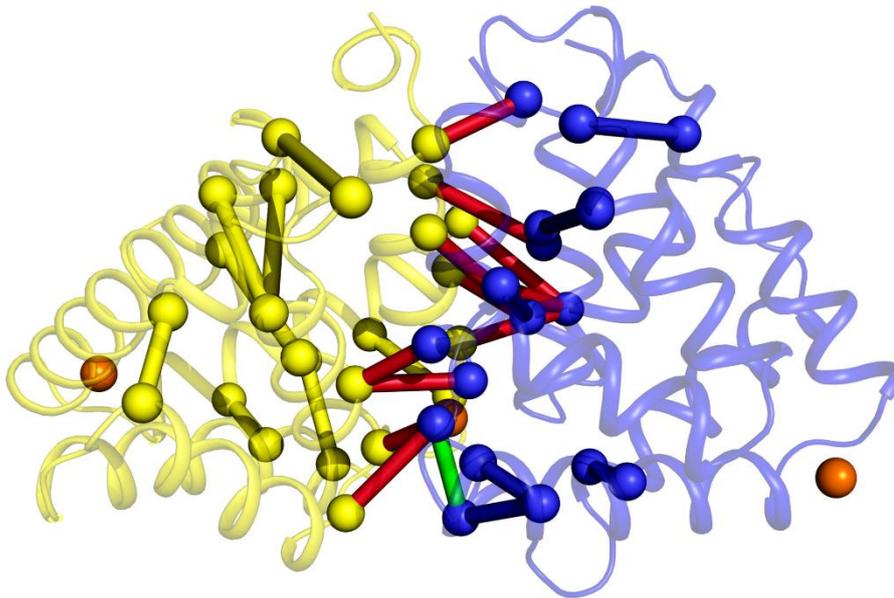


Fig 8. Overview of essential node pairs for motion transmission in Fel d 1 simulations. ‘Hinge’ pairs in each chain are colored yellow for Chain A and blue for Chain B, while interface pairs are colored red. A connection between interface and intrachain motions (detected in the system simulated with Ca^{2+} and no glycosylation) is shown in green. Calcium ions shown in orange.

doi:10.1371/journal.pone.0132311.g008

seem to have more directed movements. This is especially highlighted in the glycosylated systems, in which the chains of Fel d 1 (monomers in the fused protein) seem to move in the opposite direction of each other. Helices I and II show pronounced mobility (in chain A).

Discussion

Considering the debate on whether Ca^{2+} binds to Fel d 1 or not, for this work we employed different calculation approaches to inspect such interactions. Our results indicate that the central ion is bound to the protein, while the lateral ions are free. This may hinder the proposed action of Fel d 1 in allergy through modulation of phospholipase A_2 activity via calcium sequestration [8, 10]. This property was proposed taking uteroglobin as a model, but even for this protein there is controversy regarding the putative binding of the ion [11–13].

Supporting the concept that Ca^{2+} is not bound to the lateral sites of Fel d 1 is the physiological concentration of calcium in cats. Among the synthesis sites of Fel d 1 (which include perianal, sebaceous, and salivary glands), the calcium concentration in submandibular salivary glands (20 mmol/kg) is considered high for soft tissues in mammals (twice as much as normal human lung tissue) [23, 24], but the calcium concentration in the crystallization medium was 25-fold greater than that [8]. This would lead to unspecific binding of Ca^{2+} , and a fourth ion in the final crystallized structure is considered one of these cases [8].

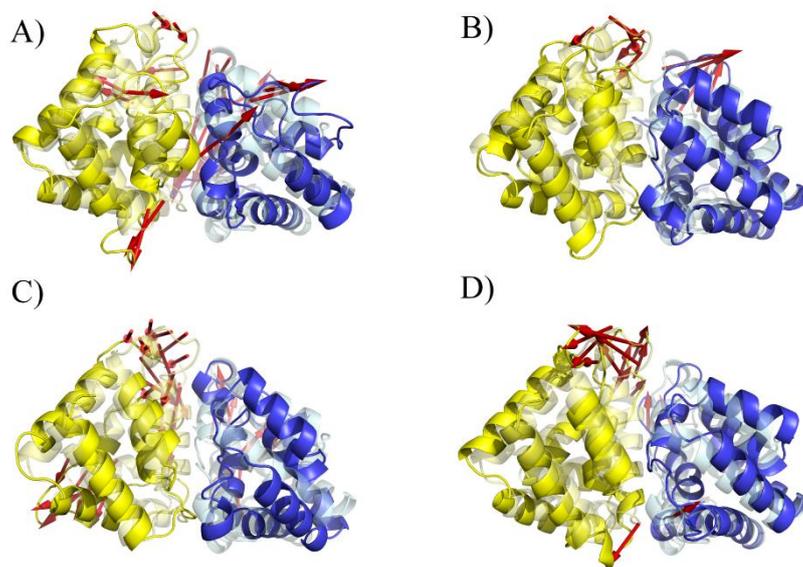


Fig 9. Overview of conformational changes observed for the different GROMOS simulated systems. The initial structures are shown as transparent, with arrows pointing the major motions observed between the initial and final stages of the MD simulations. (A) Calcium-free system; (B) Calcium-bound system; (C) Minimally glycosylated system; (D) Fully glycosylated system. Chain A shown in yellow and chain B shown in blue.

doi:10.1371/journal.pone.0132311.g009

To avoid possible parameter bias in simulating this protein-metal complex, we used three different force fields. There were differences in overall flexibility observed between all-atom force fields (AMBER and CHARMM) and united-atoms force field (GROMOS) in this work. These differences, which may be striking at first, are probably due to faster convergence and enhanced conformational sampling in the united-atoms force field [25]. The alternative explanation to GROMOS apparently not reaching convergence is that AMBER and CHARMM simulations remain stuck in energetic wells in a more rugged energy surface than in GROMOS simulations. However, the concept of convergence involves evaluating if what has been measured is enough to discard the occurrence of anything important which remains unmeasured [26]. There is no guarantee that a stable system (based on the generally used measures, such as RMSD) will remain stable if simulated for some extra nanoseconds. Besides that, convergence varies with every property analyzed and they do not correlate to each other [27]. It has been shown that evaluating convergence is a subjective matter, and that the evaluation of structural features such as RMSD for this purpose is very unreliable [28]. In the context of our work, they show that different levels of detail in the force field parametrization have no difference in describing the Ca^{2+} behavior, allowing us to compare the GROMOS results obtained for the calcium ion studies with the glycosylation studies with the same force-field, for which there is a wealth of saccharide parameters [18,19]. Besides the combination of different simulations, including the crystal environment, the semi-empirical calculations improved the evaluation of such important aspect of Fel d 1 structural biology by exploring the electronic component of

Ca^{2+} ions coordination in this complex. It is possible that Ca^{2+} ions can bind to different, new sites on Fel d 1, based on the behavior of the GROMOS simulations. This observation reinforces that, apart from the central ion binding site, the lateral binding sites are not stable and are not supported by different inspections.

Another highlighted feature of the major cat allergen is its saccharide content. The active form of Fel d 1 has two N-glycosylation sites (Fig 2), and the attached glycan chains are highly variable [14]. For this work we inspected the largest and the smallest carbohydrate chains observed linked to the protein. Different effects were observed in each case. While the conformational behavior of the fully glycosylated protein did not differ from its unglycosylated counterpart in general, the minimally glycosylated protein behaved differently. There was an increase in flexibility, due to partial unfolding of two helices of one of the Fel subunits. The fully glycosylated protein, however, presented a reduction in its cavity size, also for one of its subunits. This difference could be due to sampling limitations in our simulations, but an extension of differences between subunits detected in the crystallographic structure [8] are not discarded.

A previous study of circular dichroism on Fel d 1 observed decrease in its helical content due to removal of glycan portions [9]. In our work, we were able to pinpoint the helices that are more likely to unfold, corresponding to this previous observation. The reduction in cavity volume observed for one subunit of the fully glycosylated protein may hint to Fel d 1 glycans acting on molecular transport modulation. It has been shown previously that selective saccharide removal from glycans in human plasmatic transport proteins caused cavity alterations, affecting its binding to hormones and drugs [29].

The putative content of Fel d 1 cavities has not been identified yet [8]. A similar protein, found in the slow loris (*Nycticebus coucang*, a prosimian), is capable of causing anaphylactic shock in humans [30]. These animals are thought to use these proteins defensively, but mainly as a communication tool. Different hydrophobic compounds were identified in association with this molecule, and it is proposed that the protein structure would act as a box, containing and delivering chemical messages [31]. It has been observed that cat females secrete less Fel d 1 than males, and that males that are handling-avoidant (less frightened and more aggressive) have greater release of the allergen than handling-friendly males [32]. This may be an indication of Fel d 1 acting in information transmission in cats.

Considering the synthesis of the protein in perianal glands and its spreading by grooming involving saliva, it is tempting to propose that steroid-like molecules from the perianal glands may be entrapped in Fel d 1 by changes in pH (or other feature) brought about by interaction with saliva. Based on the crystal structure of uteroglobin bound to a steroid analog [33], we performed dockings and confirmed that the Fel d 1 cavities observed in simulation would accommodate progesterone (312 \AA^3), testosterone (399 \AA^3), and the steroid analog, PCB (335.6 \AA^3). A communication role has been proposed for the loris brachial gland protein, (a Fel d 1 fold analog) [27], and has also been proposed for Fel d 1 in the 1990s [34].

Fel d 1 is a dimer of heterodimers in its native form, and was expressed as a dimer of fused monomers for crystallization [8]. For this reason, each monomer in the crystal (originally a dimer of chains α and β) is called a chain (A and B). Despite being a homodimer, the interacting interface of the oligomer is not symmetric. Asymmetry is important for the proper recognition of each monomer (α and β dimer) in terms of immunological response [15]. The relevance of asymmetry for the biological action of the Fel d 1 dimer (or dimer of dimers in its native form) has not been investigated so far, but our results point to differential effects of glycosylation and cavity behavior in each of the asymmetrically positioned Fel d 1 subunits.

In summary, in this work we were able to apply theoretical chemical tools in the study of Fel d 1 molecular behavior and its interaction with calcium ions and glycosylation. Our results not

only corroborate previous observations, but also widen the breadth of action of the major cat allergen. We hope to have highlighted some advantages of the computational structure biology applied to the study of Fel d 1. Future investigations, focused on the influence of variable glycosylation structures on the protein, site-directed mutagenesis of metal-binding and motion-transmitting residues, and especially those focused on the functions carried out by Fel d 1 in feline physiology, may benefit from the observations presented here. For instance, the calcium-binding properties of the protein are less prominent than previously proposed, reducing the ion structural role. The protein binding cavity is compatible with previously proposed ligands, providing structural basis for these proposals. The impact of glycosylation on Fel d 1 seems limited, also reducing its importance for cavity volume control. The models presented here provide structural and rational basis to guide future studies with ligand binding and site-directed mutagenesis. We hope this kind of work may highlight the need for cooperation between laboratory experimentalists and computer modelers with the joint aim of studying structural biology in atomistic scale in both time and space dimensions.

Materials and Methods

Simulated systems

We retrieved the structure for the tetrameric form of Fel d 1 from the RCSB Protein Data Bank under PDB ID 2EJN [8], and used it to prepare nine Fel d 1 systems for MD simulations under different force fields and conditions. The crystal-derived Fel d 1 was simulated with and without crystallographic Ca^{2+} ions employing three force fields (AMBER99SB-ILDN [35], CHARMM27 [36], and GROMOS53a6 [20]). Two glycosylated forms of the protein were also simulated, as derived from MALDI-MS analyses of the native Fel d 1, basically corresponding to a desialylated (minimal or smallest) and a sialylated (full or largest) oligosaccharide structures [14]. We also regenerated the crystalline environment surrounding the tetrameric protein using symmetry tools from PyMol (Schrödinger, LLC) and symmetry group information from the Fel d 1 crystal. This procedure added ten other tetramers to this ninth simulated system.

Glycoprotein construction

Each dimer of Fel d 1 has one N-glycosylation site, at Asn103 [8], in a total of two sites for the biological tetramer [14]. We used the full length [(-Asn-) GlcNAc₂Man₃GlcNAc₃Gal₃(Fuc₁)-NeuAc₃] and truncated [(-Asn-) GlcNAc₂Man₃GlcNAc₃Gal₂] forms of the glycosidic chains (named 'full' and 'minimal' for reference) previously identified [14] to build the glycosylated proteins. This construction was made via Glycosciences.de modelling tools [37, 38]. The glycosidic linkages composing such glycans had their geometries adjusted to each of their relative abundance as isolated disaccharides in water, thus determined as their main conformational states [39, 40]. Such procedure was successfully applied in previous glycoprotein studies [41, 42]. The oligosaccharide topologies were described under GROMOS 53A6glyc parameters set [18, 19] and the protein moieties under the standard GROMOS96 53A6 force field.

Molecular dynamics studies

We submitted the Fel d 1 systems to MD simulations with the GROMACS 4.5 suite [42] for 200 ns each, with the exception of crystal-environment mimic, which we simulated for 25 ns. The systems were solvated in rectangular boxes using periodic boundary conditions, and subjected to different force fields and water models: GROMOS96 53a6 force field [20] with SPC water models [43], CHARMM27 [32] with TIP3P [44], and AMBER99SB-ILDN [31] also with TIP3P. The Lincs method [45] was applied to constrain covalent bond lengths, allowing an

integration step of 2 fs after an initial energy minimization using Steepest Descents algorithm. Electrostatic interactions were calculated with Particle Mesh Ewald method [46], in which short-range interactions cutoff values were set to 0.9 nm. Temperature and pressure were kept constant by coupling proteins, ions, and solvent to external temperature and pressure baths with coupling constants of $\tau = 0.1$ and 0.5 ps (Berendsen barostat, velocity rescaling thermostat [47, 48], respectively). The dielectric constant was treated as $\epsilon = 1$, and the reference temperature was adjusted to 300 K to reflect the temperature faced by the protein in cat hair and dandruff. The systems were slowly heated from 50 to 300 K, in steps of 5ps, each step increasing the reference temperature by 50 K, allowing a progressive equilibration of the molecular system. The simulations were performed with no restraint, with a reference value of 3.5 Å between heavy atoms for a hydrogen bond, and a cutoff angle of 30° between hydrogen-donor-acceptor [41]. The so called "crystal-environment mimic" was based on reproducing the crystallographic Fel d 1 neighboring asymmetric units only, not their entire crystallographic chemical environment. The system sizes for all simulations are listed on [S3 Table](#).

Cavities were measured with Mole2 [49, 50]. Essential nodes for motion transmission were inspected as described previously for tRNAs [51, 52]. Ligand volumes were calculated with UCSF Chimera 1.7 [53]. Docking studies of Fel d 1 chains A and B to ligands PCB, progesterone and testosterone were each performed with Autodock 4.2 [54] and lowest energy conformations for each system were analyzed with Poseview [55] and Ligplot [56]. All other analyses were carried out with dedicated tools from the Gromacs suite.

Semiempirical calculations

In order to add information on metal coordination to molecular mechanics calculations we submitted Fel d 1 with all four crystallographic Ca^{2+} ions to semiempirical quantum chemistry calculations using MOPAC [57] with the PM6 Hamiltonian [58] and MOZYME linear scaling method [59]. The following steps, based on a previous study of metal coordination [60], were employed on the calculations: 1) hydrogen atoms optimization to a gradient lower than 5 kJ/mol/Å; 2) global optimization with gradient convergence criterium of 10 kJ/mol/Å and cutoff = 6; 3) global optimization with gradient convergence criterium of 5 kJ/mol/Å and cutoff = 9. We carried out calculations in the absence of crystallographic water molecules, with the COSMO continuum solvation model [61], and with all the crystallographic water molecules within a 5 Å radius from the Ca^{2+} ions. We used the obtained minimum energy conformation to evaluate the bond orders and distances to estimate the possible coordination between amino acid residues and Ca^{2+} ions.

Supporting Information

S1 Fig. Validation of GROMOS 53A6glyc parameters implemented for sialic acid.
(PDF)

S2 Fig. Distances measured for molecular dynamics simulations of Fel d 1 under different force fields. Please refer to [Table 1](#) for details.
(PDF)

S3 Fig. Interaction energies measured between Fel d 1 monomers during molecular dynamics simulations under different force fields. Color coding: Simulations without Ca^{2+} ions (black), simulations with Ca^{2+} ions (red).
(PDF)

S4 Fig. Secondary structure content for the Fel d 1 systems simulated with GROMOS. (A) Fel d 1 without Ca^{2+} ; (B) Fel d 1 with Ca^{2+} ; (C) Fel d 1 with minimal glycosylation; (D) Fel d 1

with full glycosylation.

(PDF)

S5 Fig. All-atom root mean square deviation (RMSD) measured for Fel d 1 segments corresponding to the original chain regions (fused in the crystallized protein). (A) Calcium-free system; (B) Calcium-bound system; (C) Minimally glycosylated system; (D) Fully glycosylated system. Region A₁ in black, region A₂ in red, region B₁ in blue, region B₂ in green.

(PDF)

S6 Fig. Cavity volume overview (calcium-free system as an example). (A) initial cavity volume, (B) final cavity volume (cavity volumes shown as green surfaces); (C) cavity volume measurements for each Fel d 1 chain in four simulations.

(PDF)

S7 Fig. Structural 'oscillation' observed for (A) chain A, and (B) chain B of Fel d 1 in four different simulations. The distance between each original chain in the native Fel d 1 (half chains of the crystallographic protein) is shown. The oscillation may indicate an opening-closing behavior. (calcium-free system in black, calcium-bound system in red, minimally glycosylated system in blue, fully glycosylated system in green; distances measured for the centers of mass in each structure).

(PDF)

S1 Table. Parameters used in the GROMOS 53A6glyc force field for sialic acid, in a GRO-MACS-compatible format. These encompass atom types, atomic partial charges, charge-group definition, bond stretching, bond-angle bending, improper dihedral deformation and torsional potential dihedral.

(PDF)

S2 Table. Essential node pairs detected in four Fel d 1 simulated systems. Nodes that appear more than once are grouped by color. Interactions detected between pairs by PDBsum for the crystallographic structure are in parentheses.

(PDF)

S3 Table. Sizes of simulated systems.

(PDF)

Acknowledgments

The authors wish to thank the Centro Nacional de Supercomputação at Universidade Federal do Rio Grande do Sul for the assistance and access to the supercomputer, and the anonymous reviewers whose input greatly improved the final version of the manuscript. This work was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior), FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul), and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, grant number APQ-0398-1.06/13).

Author Contributions

Conceived and designed the experiments: RLB LGS LPF HV. Performed the experiments: RLB LGS LPF. Analyzed the data: RLB LGS. Wrote the paper: RLB LGS LPF HV.

References

- Dharmage SC, Lodge CL, Matheson MC, Campbell B, Lowe AJ. (2012) Exposure to cats: update on risks for sensitization and allergic diseases. *Curr Allergy Asthma Rep* 12:413–423. doi: [10.1007/s11882-012-0268-x](https://doi.org/10.1007/s11882-012-0268-x) PMID: [22878928](https://pubmed.ncbi.nlm.nih.gov/22878928/)
- Asher MI, Montefort S, Björkstén B, Lai CK, Strachan DP, Weiland SK, et al. (2006) Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multi country cross-sectional surveys. *Lancet* 368: 733–743. PMID: [16935684](https://pubmed.ncbi.nlm.nih.gov/16935684/)
- Kay AB. (2000) Overview of 'Allergy and allergic diseases: with a view to the future'. *Br Med Bull* 56:843–864. PMID: [11359624](https://pubmed.ncbi.nlm.nih.gov/11359624/)
- Carrer P, Maroni M, Alcini D, Cavallo D. (2001) Allergens in indoor air: environmental assessment and health effects. *Sci Total Environ* 270: 33–42. PMID: [11327396](https://pubmed.ncbi.nlm.nih.gov/11327396/)
- Svanes C, Heinrich J, Jarvis D, Chinn S, Omenaas E, Gulsvik A, et al. (2003) Pet-keeping in childhood and adult asthma and hay fever: European community respiratory health survey. *J Allergy Clin Immunol* 112: 289–300. PMID: [12897734](https://pubmed.ncbi.nlm.nih.gov/12897734/)
- Grönlund H, Saame T, Gafvelin G, van Hage M. (2010) The major cat allergen, Fel d 1, in diagnosis and therapy. *Int Arch Allergy Immunol* 151: 265–274. doi: [10.1159/000250435](https://doi.org/10.1159/000250435) PMID: [19844127](https://pubmed.ncbi.nlm.nih.gov/19844127/)
- Kaiser L, Grönlund H, Sandalova T, Ljunggren HG, van Hage-Hamsten M, Achour A, et al. (2003) The crystal structure of the major cat allergen Fel d 1, a member of the secretoglobulin family. *J Biol Chem* 278: 37730–37735. PMID: [12851385](https://pubmed.ncbi.nlm.nih.gov/12851385/)
- Kaiser L, Velickovic TC, Badia-Martinez D, Adedoyin J, Thunberg S, Hallén D, et al. (2007) Structural characterization of the tetrameric form of the major cat allergen Fel d 1. *J Mol Biol* 370: 714–727. PMID: [17543334](https://pubmed.ncbi.nlm.nih.gov/17543334/)
- Vailes LD, Li Y, Bao Y, DeGroot H, Aalberse RC, Chapman MD. (1994) Fine specificity of B-cell epitopes on *Felis domesticus* allergen I (Fel d I): effect of reduction and alkylation or deglycosylation on Fel d I structure and antibody binding. *J Allergy Clin Immunol* 93: 22–33. PMID: [7508462](https://pubmed.ncbi.nlm.nih.gov/7508462/)
- Barnes HJ, Nordlund-Møller L, Nord M, Gustafsson J, Lund, Gillner M. (1996) Structural basis for calcium binding by uteroglobins. *J Mol Biol* 256: 392–404. PMID: [8594205](https://pubmed.ncbi.nlm.nih.gov/8594205/)
- Andersson O, Nordlund-Møller L, Barnes HJ, Lund J. (1994) Heterologous expression of human uteroglobin/polychlorinated biphenyl-binding protein. Determination of ligand binding parameter and mechanism of phospholipase A2 inhibition in vitro. *J Biol Chem* 269: 19081–19087. PMID: [7518449](https://pubmed.ncbi.nlm.nih.gov/7518449/)
- Nord M, Gustafsson JA, Lund J. (1995) Calcium dependent binding of uteroglobin (PCB-BP/CCSP) to negatively charged phospholipid liposomes. *FEBS Lett* 374: 403–406. PMID: [7589580](https://pubmed.ncbi.nlm.nih.gov/7589580/)
- Chowdhury B, Mantile-Selvaggi G, Miele L, Cordella-Miele E, Zhang Z, Mukherjee AB. (2002) Lys43 and Asp46 in alpha-helix 3 of uteroglobin are essential for its phospholipase A2 inhibitory activity. *Biochem Biophys Res Commun* 295: 877–883. PMID: [12127976](https://pubmed.ncbi.nlm.nih.gov/12127976/)
- Kristensen AK, Schou C, Roepstorff P. (1997) Determination of isoforms, N-linked glycan structure and disulfide bond linkages of the major cat allergen Fel d 1 by a mass spectrometric approach. *Biol Chem* 378: 899–908. PMID: [9377487](https://pubmed.ncbi.nlm.nih.gov/9377487/)
- Seppälä U, Hägglund P, Wurtzen PA, Ipsen H, Thorsted P, Lenhard T, et al. (2005) Molecular characterization of major cat allergen Fel d 1: expression of heterodimer by use of a baculovirus expression system. *J Biol Chem* 280: 3208–3216. PMID: [15546862](https://pubmed.ncbi.nlm.nih.gov/15546862/)
- Emara M, Royer P-J, Abbas Z, Sewell HF, Mohamed GG, Singh S, et al. (2011) Recognition of the Major Cat Allergen Fel d 1 through the Cysteine-rich domain of the Mannose Receptor determines its allergenicity. *J. Biol. Chem.* 286: 13033–13040. doi: [10.1074/jbc.M111.220657](https://doi.org/10.1074/jbc.M111.220657) PMID: [21335554](https://pubmed.ncbi.nlm.nih.gov/21335554/)
- Platts-Mills TA, Woodfolk JA. (2011) Allergens and their role in the allergic immune response. *Immunol Rev* 242:51–68. doi: [10.1111/j.1600-065X.2011.01021.x](https://doi.org/10.1111/j.1600-065X.2011.01021.x) PMID: [21682738](https://pubmed.ncbi.nlm.nih.gov/21682738/)
- 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 53A6glyc parameter set for glycoproteins. *J Comput Chem* 35: 2087–2095. doi: [10.1002/jcc.23721](https://doi.org/10.1002/jcc.23721) PMID: [25196137](https://pubmed.ncbi.nlm.nih.gov/25196137/)
- 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. PMID: [15264259](https://pubmed.ncbi.nlm.nih.gov/15264259/)
- MOPAC Manual (2005) Bond order. Available at: <http://www.cache.fujitsu.com/mopac/Mopac2002manual/node464.html> (Accessed 3 December 2014).
- 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.

23. Erjavec F, Stanovnik L. (1987) Histamine and calcium secretion in the submandibular gland of the cat evoked by two types of stimulation. *Agents Actions* 20: 188–190. PMID: [2440268](#)
24. Larsson L, Öhman S. (1978) Serum ionized calcium and corrected total calcium in borderline hyperparathyroidism. *Clin Chem* 24: 1962–1965. PMID: [709830](#)
25. Hsieh MJ, Luo R. (2010) Balancing simulation accuracy and efficiency with the Amber united atom force field. *J Phys Chem B* 114: 2886–2893. doi: [10.1021/jp906701s](#) PMID: [20131885](#)
26. Grossfield A, Feller SE, Pitman MC (2007) Convergence of molecular dynamics simulations of membrane proteins. *Proteins* 67: 31–40. PMID: [17243153](#)
27. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, et al. (2006) Biomolecular modeling: Goals, problems, perspectives. *Angew Chem Int Ed Engl* 45: 4064–4092. PMID: [16761306](#)
28. Knapp B, Frantal S, Cibena M, Schreiner W, Bauer P (2011) Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *J Comput Biol* 18: 997–1005. doi: [10.1089/cmb.2010.0237](#) PMID: [21702691](#)
29. Huang RY, Hudgens JW. (2013) Effects of desialylation on human α 1-acid glycoprotein-ligand interactions. *Biochemistry* 52: 7127–7136. doi: [10.1021/bi4011094](#) PMID: [24041412](#)
30. Krane S, Itagaki Y, Nakanishi K, Weldon PJ. (2003) "Venom" of the slow loris: sequence similarity of prosimian skin gland protein and Fel d 1 cat allergen. *Naturwissenschaften* 90: 60–62. PMID: [12590298](#)
31. Hagey LR, Fry BG, Fitch-Snyder H. (2007) Talking defensively: a dualuse for the brachial gland exudate of slow and pygmy lorises. In: Gurski S, Nekarai KAI, editors. *Primate Anti-predatory Strategies*, vol. 2. New York: Springer, pp. 253–272.
32. Bienboire-Frosini C, Cozzi A, Lafont-Lecuelle C, Vervloet D, Ronin C, Pageat P. (2012) Immunological differences in the global release of the major cat allergen Fel d 1 are influenced by sex and behaviour. *Vet J* 193: 162–167. doi: [10.1016/j.tvjl.2011.09.031](#) PMID: [22044714](#)
33. Hård T, Barnes HJ, Larsson C, Gustafsson JA, Lund J. (1995) Solution structure of a mammalian PCB-binding protein in complex with a PCB. *Nat Struct Biol*. 2: 983–989. PMID: [7583672](#)
34. Dornelas de Andrade AD, Birnbaum J, Magalon C, Magnol JP, Lanteaume A, Charpin D, et al. (1996) Fel d I levels in cat anal glands. *Clin Exp Allergy* 26: 178–180. PMID: [8835125](#)
35. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78: 1950–1958. doi: [10.1002/prot.22711](#) PMID: [20408171](#)
36. MacKerell AD Jr., Banavali N, Foloppe N. (2001) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56: 257–265.
37. Lütke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth CW. (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology* 16: 71R–81R. PMID: [16239495](#)
38. Frank M, Schloissnig S. (2010) Bioinformatics and molecular modeling in glycobiology. *Cell Mol Life Sci* 67: 2749–2772. doi: [10.1007/s00018-010-0352-4](#) PMID: [20364395](#)
39. Fernandes CL, Sachett LG, Pol-Fachin L, Verli H. (2009) GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins. *Carbohydr Res* 345: 663–671. doi: [10.1016/j.carres.2009.12.018](#) PMID: [20106471](#)
40. Pol-Fachin L, Becker CF, Guimarães JA, Verli H. (2011) Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Proteins* 29: 2735–2745.
41. Virgens MYF, Pol-Fachin L, Verli H, Saraiva-Pereira ML. (2014) Effects of glycosylation and pH conditions in the dynamics of human arylsulfatase A. *J Biomol Struct Dyn* 32: 567–579. doi: [10.1080/07391102.2013.780982](#) PMID: [23581857](#)
42. Hess B, Kutzner C, van der Spoel D, Lindahl E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4: 435–447.
43. Berendsen HJC, Grigera JR, Straatsma TP. (1987) The Missing Term in Effective Pair Potentials. *J Phys Chem* 91: 6269–6271.
44. Jorgensen WL, Madura JD. (1983) Solvation and conformation of methanol in water. *J Am Chem Soc* 105: 1407–1413.
45. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18: 1463–1472.
46. 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.

47. Berendsen HJC, Postma JPM, DiNola A, Haak JR. (1984) Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81: 3684–3690.
48. Bussi G, Donadio D, Parrinello M. (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101. PMID: [17212484](#)
49. Berka K, Hanak O, Sehnal D, Banas P Navratilova V, Jaiswal D, et al. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res* 40: W222–W227. doi: [10.1093/nar/gks363](#) PMID: [22553366](#)
50. Sehnal D, Svobodova Varekova R, Berka K, Pravda L, Navratilova V, Banas P, et al. (2013) MOLE 2.0: Advanced approach for analysis of biomacromolecular channels. *J Cheminform* 5: 39. doi: [10.1186/1758-2946-5-39](#) PMID: [23953065](#)
51. Sethi A, Eargle J, Black AA, Luthy-Schulten Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A*. 106:6620–6625. doi: [10.1073/pnas.0810961106](#) PMID: [19351898](#)
52. Black Pyrkosz A, Eargle J, Sethi A, Luthy-Schulten Z. (2010) Exit strategies for charged tRNA from GluRS. *J Mol Biol* 16:1350–1371.
53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004) *J Comput Chem* 25:1605–1612. PMID: [15264254](#)
54. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS et al. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry* 16: 2785–2791.
55. Stierand K, Maaß P, Rarey M. (2006) Molecular Complexes at a Glance: Automated Generation of two-dimensional Complex Diagrams. *Bioinformatics*, 22: 1710–1716. PMID: [16632493](#)
56. Wallace AC, Laskowski RA, Thornton JM. (1996) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8: 127–134.
57. Stewart J. (2008) MOPAC2009, Stewart Computational Chemistry, Colorado Springs, USA, Retrieved from <http://OpenMOPAC.net>.
58. Stewart JP. (2007). Optimization of parameters for semi empirical methods V: Modification of nndc approximations and application to 70 elements. *J Mol Model* 13: 1173–1213. PMID: [17828561](#)
59. Stewart JP. (1996). Application of localized molecular orbitals to the solution of semi empirical self-consistent field equations. *Int J Quant Chem* 58: 133–146.
60. Fernandes CL, Escouto GB, Verli H. (2014) Structural glyco-biology of heparinase II from *Pedobacter-heparinus*. *J Biomol Struct Dyn* 32:1092–1102. doi: [10.1080/07391102.2013.809604](#) PMID: [23808670](#)
61. Klamt A, Schüürmann G. (1993) COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc, Perkin Trans 2*: 799–805.

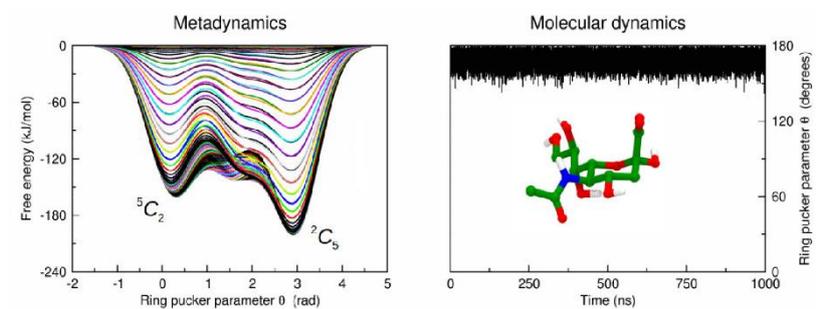


Fig. S1. Validation of GROMOS 53A6glyc parameters implemented for sialic acid.

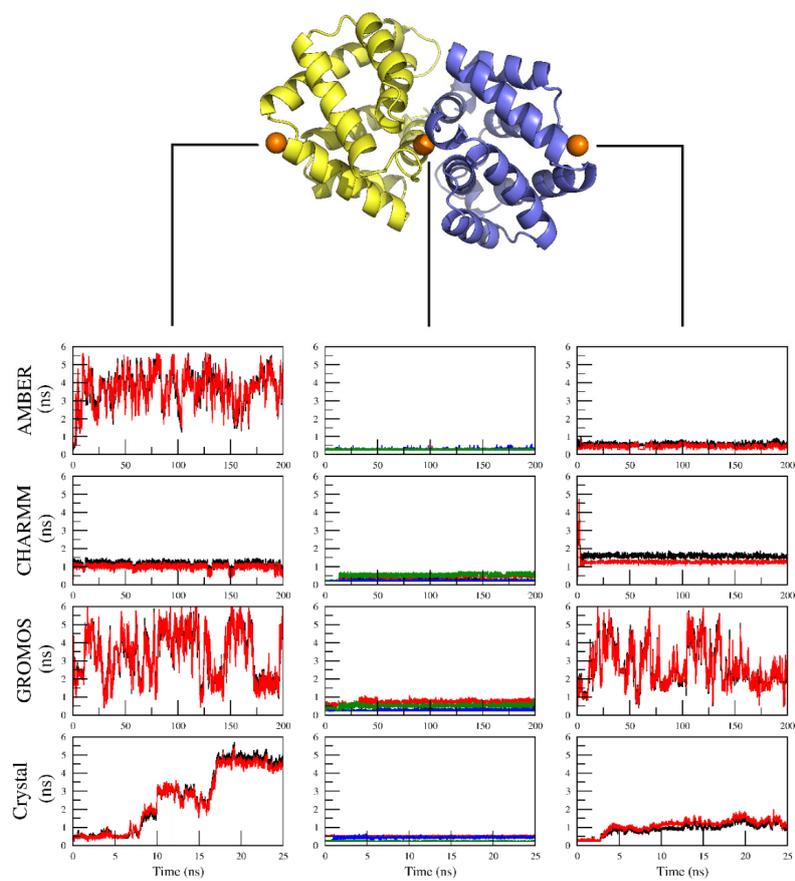


Fig. S2. Distances measured for molecular dynamics simulations of Fel d 1 under different force fields. Please refer to Table 1 for details.

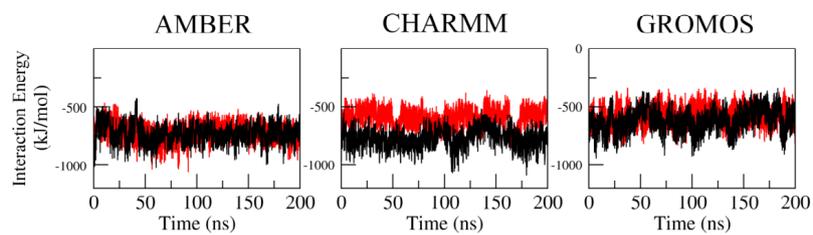


Fig. S3. Interaction energies measured between Fel d 1 monomers during molecular dynamics simulations under different force fields. Color coding: Simulations without Ca²⁺ ions (black), simulations with Ca²⁺ ions (red).

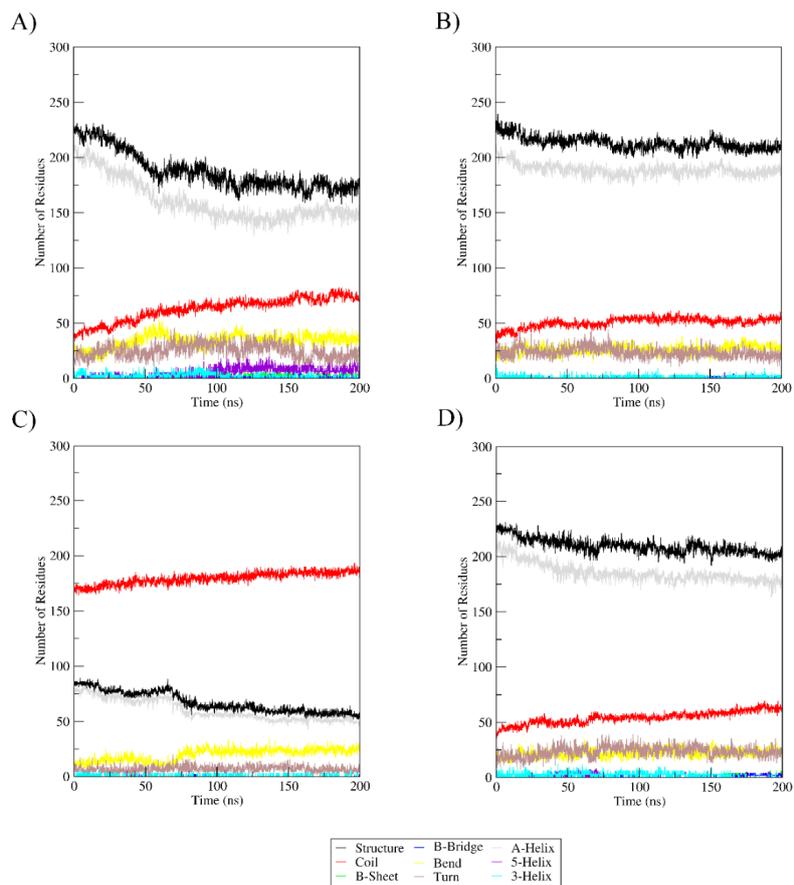


Fig. S4. Secondary structure content for the Fel d 1 systems simulated with GROMOS. (A) Fel d 1 without Ca^{2+} ; (B) Fel d 1 with Ca^{2+} ; (C) Fel d 1 with minimal glycosylation; (D) Fel d 1 with full glycosylation.

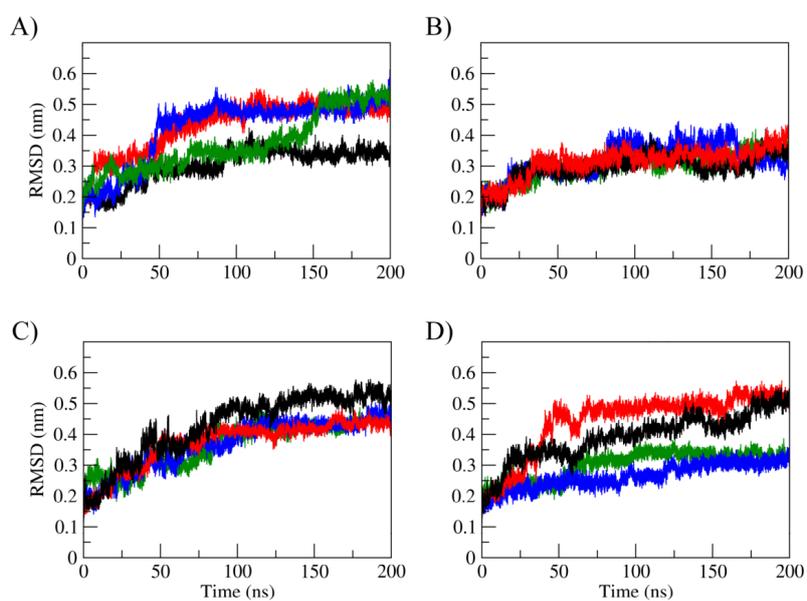


Fig. S5. All-atom root mean square deviation (RMSD) measured for Fel d 1 segments corresponding to the original chain regions (fused in the crystallized protein). (A) Calcium-free system; (B) Calcium-bound system; (C) Minimally glycosylated system; (D) Fully glycosylated system. Region A₁ in black, region A₂ in red, region B₁ in blue, region B₂ in green.

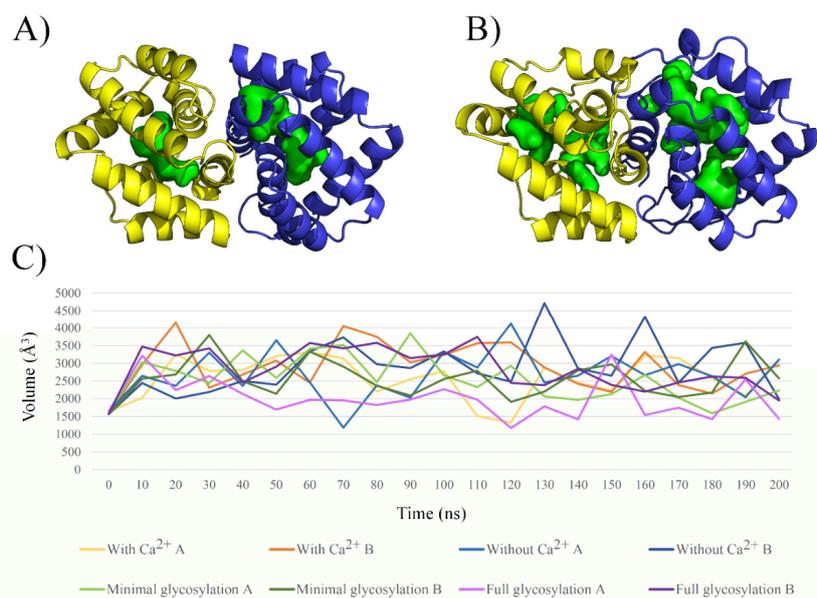


Fig. S6. Cavity volume overview (calcium-free system as an example). (A) initial cavity volume, (B) final cavity volume (cavity volumes shown as green surfaces); (C) cavity volume measurements for each Fel d 1 chain in four simulations.

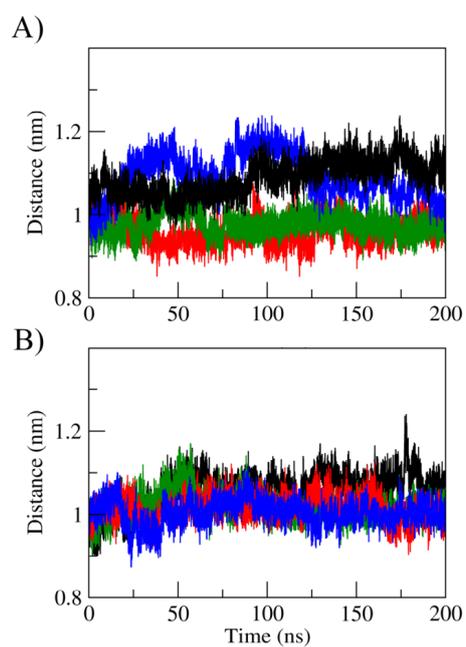


Fig. S7. Structural ‘oscillation’ observed for (A) chain A, and (B) chain B of Fel d 1 in four different simulations. The distance between each original chain in the native Fel d 1 (half chains of the crystallographic protein) is shown. The oscillation may indicate an opening-closing behavior. (calcium-free system in black, calcium-bound system in red, minimally glycosylated system in blue, fully glycosylated system in green; distances measured for the centers of mass in each structure).

Table S1. Parameters used in the GROMOS 53A6glyc force field for sialic acid, in a GROMACS-compatible format. These encompass atom types, atomic partial charges, charge-group definition, bond stretching, bond-angle bending, improper dihedral deformation and torsional potential dihedral.

```
[ NEUN ]
[ atoms ]
C5 CH1 0.00000 0
N5 N -0.28000 0
HN5 H 0.28000 0
C5A C 0.38000 0
O5A O -0.38000 0
C5B CH3 0.00000 0
C4 CH1 0.23200 1
O4 OA -0.64200 1
HO4 H 0.41000 1
C7 CH1 0.23200 2
O7 OA -0.64200 2
HO7 H 0.41000 2
C8 CH1 0.23200 3
O8 OA -0.64200 3
HO8 H 0.41000 3
C9 CH2 0.23200 4
O9 OA -0.64200 4
HO9 H 0.41000 4
C6 CH1 0.37600 5
O6 OA -0.48000 5
C2 CH1 0.23200 5
O2 OA -0.53800 5
HO2 H 0.41000 5
C3 CH2 0.00000 5
C1 C 0.36000 6
O1A OM -0.68000 6
O1B OM -0.68000 6

[ bonds ]
C5 N5 gb_21
N5 HN5 gb_2
N5 C5A gb_10
C5A O5A gb_5
C5A C5B gb_16
C5 C4 gb_26
C5 C6 gb_26
C4 O4 gb_20
C4 C3 gb_26
O4 HO4 gb_1
C3 C2 gb_26
C7 C6 gb_26
C7 O7 gb_20
O7 HO7 gb_1
C7 C8 gb_26
C8 O8 gb_20
O8 HO8 gb_1
C9 C8 gb_26
C9 O9 gb_20
O9 HO9 gb_1
C6 O6 gb_20
O6 C2 gb_20
C2 O2 gb_20
O2 HO2 gb_1
C1 O1A gb_6
C1 O1B gb_6
C1 C2 gb_26

[ angles ]
; ai aj ak gromos type
C6 C5 N5 ga_15
N5 C5 C4 ga_15
C5 N5 HN5 ga_18
C5 N5 C5A ga_31
HN5 N5 C5A ga_32
N5 C5A O5A ga_33
N5 C5A C5B ga_19
O5A C5A C5B ga_30
C4 C5 C6 ga_8
C5 C4 O4 ga_9
C5 C4 C3 ga_8
O4 C4 C3 ga_9
C4 O4 HO4 ga_12
C4 C3 C2 ga_8
C5 C6 C7 ga_8

C5 C6 O6 ga_9
C7 C6 O6 ga_9
HO7 O7 C7 ga_12
O7 C7 C6 ga_9
O7 C7 C8 ga_9
C6 C7 C8 ga_8
HO8 O8 C8 ga_12
O8 C8 C7 ga_9
O8 C8 C9 ga_9
C7 C8 C9 ga_8
O9 C9 C8 ga_9
HO9 O9 C9 ga_12
C6 O6 C2 ga_10
C3 C2 O6 ga_9
C3 C2 O2 ga_9
O6 C2 O2 ga_9
O6 C2 C1 ga_9
O2 C2 C1 ga_9
C2 O2 HO2 ga_12
O1A C1 O1B ga_38
O1A C1 C2 ga_22
O1B C1 C2 ga_22
C3 C2 C1 ga_8

[ impropers ]
; ai aj ak al gromos type
N5 C5A C5B HN5 gi_1
C5A C5B N5 O5A gi_1
C5 O6 C7 C6 gi_2
C5 O4 C3 C4 gi_2
C5 N5 C4 C6 gi_2
C2 O6 O2 C3 gi_2
C3 O6 C1 C2 gi_2
C1 O1A O1B C2 gi_1

[ dihedrals ]
; ai aj ak al gromos type
C3 C4 C5 N5 gd_16
C7 C6 C5 N5 gd_16
O4 C4 C5 N5 gd_17
O6 C6 C5 N5 gd_18
C4 C5 N5 C5A gd_39
C5 N5 C5A C5B gd_14
C6 C5 C4 O4 gd_43
C6 C5 C4 O4 gd_44
C6 C5 C4 C3 gd_34
C4 C5 C6 C7 gd_34
C4 C5 C6 O6 gd_43
C4 C5 C6 O6 gd_44
C3 C4 O4 HO4 gd_30
C5 C4 C3 C2 gd_34
O4 C4 C3 C2 gd_44
C4 C3 C2 O6 gd_44
C4 C3 C2 O2 gd_43
C4 C3 C2 O2 gd_44
C6 C7 O7 HO7 gd_30
O7 C7 C6 O6 gd_5
O7 C7 C6 O6 gd_37
C6 C7 C8 O8 gd_43
C6 C7 C8 O8 gd_44
O7 C7 C8 O8 gd_18
C7 C8 O8 HO8 gd_30
C7 C8 C9 O9 gd_43
C7 C8 C9 O9 gd_44
O8 C8 C9 O9 gd_18
C8 C9 O9 HO9 gd_30
O1A C1 C2 C3 gd_40
C5 C6 O6 C2 gd_29
C6 O6 C2 C3 gd_29
O6 C2 O2 HO2 gd_2
O6 C2 O2 HO2 gd_32

[ exclusions ]
HO2 O6
```

Table S2. Essential node pairs detected in four Fel d 1 simulated systems. Nodes that appear more than once are grouped by color. Interactions detected between pairs by PDBsum for the crystallographic structure are in parentheses.

	Interface (Chain A - Chain B)	Chain A	Chain B
Without Ca ²⁺	Lys29-Asp82 (1HB, NBC)	Cys44-Cys118 (DS)	Leu41-Tyr119 (NBC)
	Leu132-Leu94	Ile64-Val102	Phe85-Met134 (NBC)
With Ca ²⁺	Leu31-Leu94	Cys3-Ala74	Val6-Tyr81
	Asn91-Leu129 (NBC)	Val10-Phe84	Asn37-Arg127
	Leu132-Phe85 (NBC)		
Minimal Glycosylation	Phe85-Asp130 (NBC)	Val6-Phe85	Val6-Phe85
	Val128-Leu94	Ile40-Asn122 (NBC)	Asn37-Arg127
	Ser139-Thr76 (NBC)		
Full Glycosylation	Phe85-Leu132 (NBC)	Lys7-Phe80	Cys3-Tyr65
	Asn89-Leu129 (NBC)	Asn37-Arg127 (2HB)	Asn37-Val128
	Thr135-Phe85 (NBC)		

HB: hydrogen bond(s), NBC: non-bonded contacts, DS: disulfide bridge.

Note: There is a pair in the simulation with Ca²⁺ (Leu129B-Arg127B) that connects the interface pair (Asn91-Leu129) to the 'hinge' pair (Arg127-Asn37).

Table S3. Sizes of simulated systems.

System		Number of atoms	Number of counterions (Na ⁺)
AMBER	With Ca ²⁺	28,951	8
	Without Ca ²⁺	31,382	16
CHARMM	With Ca ²⁺	31,750	8
	Without Ca ²⁺	31,382	16
GROMOS	With Ca ²⁺	34,627	8
	Without Ca ²⁺	28,559	16
Crystal mimic	With Ca ²⁺	586,592	88
Glycosylated	Full glycosylation	88,194	14
	Minimal glycosylation	65,078	8

4.3 Manuscrito: Conformational characterization of fucosyltransferase 8 (FUT8)
in solution

Liana Guimarães Sachett, Eric Allison Philot, Luis Paulo Barbour Scott and Hugo Verli.

Este trabalho será submetido à revista “Jornal of Molecular Graphics and Modelling”.

Conformational characterization of fucosyltransferase 8 (FUT8) in solution

Liana Guimarães Sachett^a, Eric Allison Philot^b, Luis Paulo Barbour Scott^b and Hugo Verli^{a*}.

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

^b*Laboratório de Biologia Computacional e Bioinformática, Universidade Federal do ABC, Santo André, SP, Brazil*

Abstract

Fucosyltransferase 8 (FUT8) is a α -1,6 fucosyltransferase important for physiological and pathological processes in mammals. This enzyme is found in high expression levels in lung-cancer, gastric-cancer and colon-cancer cell lines. While terminal fucosylation, which is synthesized by FUT1–FUT7 and FUT9, is directly associated with tumor metastasis, the biological significance of core fucosylation, α 1,6-fucosylation, remains almost unknown. Here we present a molecular dynamics study on FUT8 comparing the Apo and Holo forms of the enzyme, the presence and absence of a small loop region in the catalytic site and its interaction to a target N-glycan and a target glycoprotein, integrin α 5 β 1.

Keywords:

Fucosyltransferase 8, molecular dynamics, core fucosylation

Abbreviations:

FUT8: fucosyltransferase 8; GlcNAc: N-acetylglucosamine; MD: molecular dynamics;

Introduction

Fucosyltransferase 8 (FUT8) is an eukaryotic α -1,6 fucosyltransferase catalyzing the transference of a fucose residue from GDP- β -L-fucose to the reducing terminal N-acetylglucosamine (GlcNAc) of the core structure of N-glycans (core fucosylation). This type of fucosylation is frequently observed in N-glycans of α -fetoprotein, a well-known tumor marker for hepatocellular carcinoma. Core fucosylation of N-glycans in human immunoglobulin (Ig) G1 regulates antibody-dependent cellular toxicity (ADCC) (Shields et al., 2002). The lack of core fucose on IgG1 results in an increase in ADCC activity up to 100 fold, therefore suggesting that core fucose-deficient IgG1 would be useful in antibody therapy in cancer treatment (Shinkawa et al., 2003). While terminal fucosylation, which is synthesized by FUT1–FUT7 and FUT9, is directly associated with tumor metastasis (because it is associated with the formation of sialyl-Lewis X/Lewis X and sialyl-Lewis A/Lewis X), the biological significance of core fucosylation, α 1,6-fucosylation, remains almost unknown (Listinsky et al., 2011).

FUT8 is found in high expression levels in lung-cancer, gastric-cancer and colon-cancer cell lines (Miyoshi et al, 1997). The increased expression and activity of FUT8 in the liver are strongly linked to the age-related changes in glycosylation and increased FUT8 and fucosylation influence IGF-1 signaling, demonstrating that the glycosylation machinery in liver cells is significantly affected during aging and that age-related increased FUT8 activity could influence the aging process by altering the sensitivity of the IGF-1R signaling pathway (Vanhoorem et al., 2011). Knockout studies have also shown that the lack of *FUT8* gene in mice can

lead to emphysema-like (Wang et al., 2009) and schizophrenia-like phenotypes (Fukuda et al., 2011).

The overall structure of FUT8 is comprised of two long antiparallel α -helices (residues 109–173, α 1 and α 2 helices), which form a coiled-coil structure at the N-terminus; a putative catalytic domain, comprised of two structures, an open sheet α/β structure and a Rossmann fold, which is frequently found in nucleotide binding proteins including glycosyltransferases; and a SH3 domain, which has been reported in various types of cytosolic proteins, located at the C-terminus of FUT8 (Ihara et al., 2007). A model for binding of the substrate, GDP- β -L-fucose, has recently been proposed (Kötzler et al., 2012), but a comparison between the Apo enzyme and the complexed enzyme, in solution, has not been carried out so far. Here we present a molecular dynamics study of this enzyme, free in solution and complexed to its substrate, as well as a comparison between FUT8 in the presence and in the absence of a five amino-acid disordered region, which may be important for positioning the substrate. Molecular docking studies on binding of FUT8 to its target glycan and to a target glycoprotein are also presented.

Material and Methods

Systems

We studied the behavior of FUT8 in its free form and complexed to its substrate GDP- β -L-fucose. Considering the importance of flexible loops in the catalytic mechanism of other glycosyltransferases (Ihara et al., 2007), the enzyme was studied with and without a five amino-acid disordered region (residues 368 to

372) that is predicted to be a flexible loop. This region is absent in the crystallographic structure for human FUT8, which was retrieved from Protein Data Bank (PDB ID 2DE0). We modeled the loop region using MODELLER (Fiser et al., 2000). The structure of GDP- β -L-fucose was obtained from PDB 3ZY5, a crystallographic structure of an α -1,2 fucosyltransferase and placed into the FUT8 catalytic site by superimposition of the catalytic region. Although FUT8 does not present homology to any other fucosyltransferase, there are three small regions that are highly conserved among α -1,2 and α -1,6 fucosyltransferase and that participate in the binding site for the substrate (Ihara et al., 2007). Furthermore, site directed mutagenesis studies indicate that two arginine residues in one of the conserved regions, Arg 365 e 366 in human FUT8 play important roles in donor binding.

Molecular dynamics studies

All the systems (FUT8 in its free form and complexed to GDP- β -L-fucose, with and without the flexible loop) were submitted to molecular dynamics (MD) studies using the GROMACS 4.5.1 (Hess et al., 2008) simulation suite. We studied FUT8 in its free form using Gromos96 53a6 (Oostenbrink et al., 2004) and AMBER99SB-ILDN force field (Lindorff-Larsen et al., 2010). FUT8 complexed to its substrate was studied only with Gromos96 53a6 force field, for which we have previously validated parameters for fucose (Pol-Fachin et al., 2012). Systems were placed in a triclinic box using the appropriate water model for each force field (SPC for Gromos96 53a6 and Tip3P for AMBER). Proper counter ions were used when needed. To constrain covalent bond lengths, the LINCS method (Hess et al., 1997) was applied, therefore allowing an integration step of 2fs after a first

energy minimization by using the steepest descent algorithm. The particle mesh Ewald method (Darden et al., 1993) was used in the calculation of electrostatic interactions. Temperature and pressure were kept constant by coupling the protein, ligands, ions, and solvent to external temperature and pressure baths with coupling constants of $\tau = 0.1$ and 0.5 ps (Berendsen, 1984), respectively. Parameters for GDP- β -L-fucose were built combining ATP and guanosine parameters available in the Gromos96 53a6 force field (Oostenbrink et al., 2004) and fucose parameters previously validated (Pol-Fachin et al., 2012).

Normal modes analysis

We used Normal Modes Analysis (NMA) to investigate the flexibility of the models of FUT8. We performed all calculations with the CHARMM package (Brooks et al., 1983) and we used CHARMM-GUI web server (Jo et al., 2008) to prepare the input files. In the first step, we minimized the initial structures using the parameter set PARAM22 (MacKerell et al., 1998). The minimization protocol combined Conjugate Gradient (CG) and Adopted Basis Newton-Raphson (ABNR). First, we applied the CG with harmonic constraints being progressively decreased from 250 to 0 kcal mol⁻¹ Å⁻². In a second step, we used ABNR method without constraints and with a root mean square energy gradient of 10^{-6} kcal mol⁻¹ Å⁻¹ convergence criteria. After minimization, we calculated the 60 lowest frequency normal modes using the DIMB method (Perahia and Mouawad, 1995) and the Root Mean Square Fluctuations (RMSF) of C α atoms were calculated from these modes. The non-bonded interactions were treated using a switch function being applied between 8.0 and 10.0 Å and a distance dependent dielectric function (Louet et al., 2013).

Molecular docking studies

The enzyme complex FUT8-GDP- β -L-fucose was docked to a target N-glycan ((Asn-)GlcNAc₂-Man₃-GlcNAc₇) using Autodock 4.2 (Morris et al., 2009). In this study, the substrate-enzyme complex was maintained in a rigid conformation and the N-glycan was used in a conformation previously obtained in solution. The resulting complex was further studied by MD simulation, using the Gromos96 53a6 force field as described above.

FUT8 was also docked to a target protein - the headpiece of integrin α 5 β 1, (PDB ID 3VI3) - using PIPER (Kozakov et al., 2006) via ClusPro 2.0 (Comeau et al., 2004), a web-based program validated for prediction of protein-protein interactions. It is known that about 50% of the integrin N-linked saccharides are core fucosylated (Litynska et al., 2002; Nakagawa et al., 1996). Therefore, we selected the N-glycosylation sites (present in chains A, B, C and D) from the integrin for the docking studies.

Results

Protein stability

The root mean square deviation (RMSD) and dictionary of secondary structure of proteins (DSSP) both revealed that the N-terminal portion (the coiled-coil domain) partially unfolds in all simulations run with the Gromos force field. This does not happen with the Amber force field, although this domain flexibility is similar in both force fields (Figure 1). The secondary structure content in the Gromos force field is of approximately 36% alpha-helix, 14% beta-sheet and 24% coil for the Apo

system; 34% alpha-helix, 13,5% beta-sheet and 25% coil for the complexed system; and for the Amber force field the content is around 38,5% alpha-helix, 14% beta-sheet and 20% coil (average values for all systems).

Systems flexibility

The Root Mean Square Fluctuation (RMSF) and Principal Component Analysis (PCA) revealed that, although the stability of the systems were larger in the Amber force field, the general flexibility for the Amber force field was similar to the Gromos force field, and even larger for the N-terminal domain in the Amber force-field (Figure 2). The observed flexibility is focused on the first principal component (PCA-Vector 1), and therefore in one of the principal components of flexibility. This indicates that the coiled-coil movement is intense in either force field. NMA also revealed a large N-terminal fluctuation in all systems (Supplementary data – Figure S1).

Substrate complexation

In order to substantiate the role of a five amino acid residues loop within the catalytic site of FUT8, we compared the substrate orientation within its site in the presence and in the absence of these residues. After a 100ns MD study we were able to observe that the substrate RMSD was smaller in the presence of the residues (Figure 3a). The visualization of the initial (0ns) and final orientation (100ns) of GFB within the catalytic site after the MD (Figure 3b and c) also showed that the presence of the loop helps to maintain the substrate in its initial position.

Target glycan docking to FUT8

The N-glycan-FUT8 docking resulted in a complex in agreement with the catalytic mechanism (Figure 4). Ligplot analysis performed by PDBsum server stated that the N-glycan interacts with FUT8 by hydrophobic contacts with 10 residues (Cys222, Phe467, Ser469, Arg475, Tyr497, Tyr498, Ala532, Gly533, Val543, Thr550) and hydrogen bonds with 6 other residues (Gly217, Cys218, Gln470, Asp495, Ile496 and Lys541; all hydrogen bonds with GlcNAc residues). This obtained complex demonstrated to be stable during the MD study carried out in solution (Supplementary data – Figure S2).

Molecular docking of FUT8 to integrin $\alpha 5\beta 1$

The molecular docking results obtained from ClusPro retrieved several models. The models revealed interaction complexes of FUT8 and Integrin compatible to core fucosylation of the Integrin N-glycans. We choose a model compatible to the fucosylation of chain A of the Integrin (at Asn275) to perform deeper analysis. In this model, we were able to observe that FUT8 interacts with three chains of the Integrin: chains A, C and D (Figure 5). The majority of the interactions are with chain A (13 hydrogen bonds and 143 non-bonded contacts), followed by chain D (2 hydrogen bonds and 43 non-bonded contacts) and chain C (11 non-bonded contacts). Residues from FUT8 that participate in hydrogen bonds with chain A are Lys216, His298, Gln470, Gly501, Ala518, Asp519, Arg545, Lys546, Arg549 and Thr550.

Discussion

The molecular dynamics studies presented in this work revealed that the N-terminal domain of FUT8 (coiled-coil domain) is extremely flexible when placed in solution. This flexibility is not dependent on the force field or absence/presence of the substrate, as demonstrated by MD with AMBER and Gromos force field and NMA with CHARMM. Although the membrane portion of the enzyme is absent in the structure (PDB 2DE0 starts on residue 108), it corresponds only to the first 29 amino acids. Residues 30 to 107 are predicted to be coils and short α -helices (predictions made with PHYRE - Kelley and Sternberg, 2009 - and PSIPRED – Jones, 1999 - secondary prediction servers). Therefore, we believe that this movement observed during the MD occurs when the enzyme is attached to the membrane.

We were also able to observe that the five amino acid loop region absent in the crystallographic structure is important for maintaining the substrate conformation and orientation within the putative catalytic site. This indicates that this loop may be important for the catalytic mechanism, like occurs in other GTs (Ihara et al., 2007).

Molecular docking studies retrieved an N-glycan-FUT8 complex compatible to the catalytic reaction and presented an insight into the acceptor-binding site of this enzyme, comprised by residues Cys222, Phe467, Ser469, Arg475, Tyr497, Tyr498, Ala532, Gly533, Val543, Thr550 (hydrophobic contacts) and Gly217, Cys218, Gln470, Asp495, Ile496 and Lys541 (hydrogen bonds with GlcNAc residues). This complex demonstrated to be stable during MD.

Molecular docking of Integrin $\alpha 5\beta 1$ to FUT8 resulted in a complex compatible to fucosylation of the N-glycan at Asn275 from chain A of the Integrin. 10 residues from FUT8 are hydrogen bonded with Integrin $\alpha 5\beta 1$ -chain A: Lys216, His298, Gln470, Gly501, Ala518, Asp519, Arg545, Lys546, Arg549 and Thr550. Other than that, several residues participate in hydrophobic contacts: Lys216, Asp295, Leu297, His298, Arg300, Gln470, Tyr498, Phe499, Gly500, Gly501, Pro515, Ala518, Asp519, Glu520, Ile521, Pro522, Lys541, Val543, Arg545, Lys546, Leu547, Gly548, Arg549, Thr550. Therefore, docking of Integrin $\alpha 5\beta 1$ provided an insight into FUT8 interaction region with a target protein.

Conclusions

The conformation study of FUT8 presented here demonstrated the high flexibility of the coiled-coil domain, regardless of the force field and/or presence of the substrate. This result was supported by molecular dynamics studies and normal modes analysis. We were also able to verify that the five amino acid loop region absent from the crystallographic structure acts in the catalytic site maintained the substrate orientation. Our docking studies resulted on complexes compatible to core fucosylation of an N-glycan and core fucosylation of a target glycoprotein, integrin $\alpha 5\beta 1$, and provide insights into the FUT8 binding sites.

References:

Berendsen HJC, Postma JPM, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984, 81:3684–3690.

Berendsen HJC, Grigera JR, Straatsma TP. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* 1987, 91:6269–6271.

Brooks C, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA* 1983, 80:6571–6575.

Comeau SR, Gatchell DW, Vajda S, Camacho, CJ. ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* 2004, 32:W96–W99.

Darden T, York D, Pedersen L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 1993, 98:10089–10092.

Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. Comparative Protein Structure Modeling With MODELLER. *Curr. Protoc. Bioinformatics*, John Wiley & Sons, Inc., 2006, Supplement 15, 5.6.1-5.6.30.

Fiser A, Do RK, Sali A. Modeling of loops in protein structures, *Protein Science* 2000, 9:1753-1773.

Fukuda T, Hashimoto H, Okayasu N, Kameyama A, Onogi H, Nakagawasai O, Nakazawa T, Kurosawa T, Hao Y, Isaji T, Tadano T, Narimatsu H, Taniguchi N, Gu J. 1,6-Fucosyltransferase-deficient Mice Exhibit Multiple Behavioral Abnormalities Associated with a Schizophrenia-like Phenotype. *J Biol Chem.*, 2011, 286:18434–18443.

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997, 18:1463–1472.

Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 2008, 4:435–447.

Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, Nakagawa A, and Taniguchi N. Crystal structure of mammalian α 1,6-fucosyltransferase, FUT8. *Glycobiology*, 2007, 17:455-466.

Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics.* 2006, 65:392–406.

Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 1999, 292:195-202.

Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* 2008, 29:1859-1865.

Kelley LA and Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 2009, 4:363-371.

Kötzler MP, Blank S, Bantleon FI, Spillner E, Meyer B. Donor substrate binding and enzymatic mechanism of human core α 1,6-fucosyltransferase (FUT8). *Biochim. Biophys. Acta* 2012, 1820:1915-1925.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010, 78:1950–1958.

Listinsky JJ, Siegal GP, Listinsky CM. The emerging importance of α -L-fucose in human breast cancer: a review. *Am J Transl Res* 2011, 3:292-322.

Litynska A, Przybylo M, Pochec E, Laidler P. Adhesion properties of human bladder cell lines with extracellular matrix components: the role of integrins and glycosylation. *Acta Biochim. Pol.* 2002, 49:643–650.

Louet M, Karakas E, Perret A, Perahia D, Martinez J, Floquet N. Conformational restriction of G-proteins Coupled Receptors (GPCRs) upon complexation to G-proteins: A putative activation mode of GPCRs? *FEBS Letters* 2013, 587:2656–2661.

MacKerell AD, Bashford D, Bellott M, Dunbrack, RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 1998, 102:3586–3616.

Miyoshi E, Uozumi N, Noda K, Hayashi N, Hori M, Taniguchi N. Expression of α 1-6 Fucosyltransferase in rat tissues and human cancer cell lines. *Int. J. Cancer* 1997, 72:1117–1121.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009, 16:2785-2791.

Nakagawa H, Zheng M, Hakomori S, Tsukamoto Y, Kawamura Y, Takahashi N. Detailed Oligosaccharide Structures of Human Integrin α 5 β 1 Analyzed by a Three-Dimensional Mapping Technique. *Eur. J. Biochem.* 1996, 237:76–85.

Oostenbrink C, Villa A, Mark AE, van Gunsteren W. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 2004, 25:1656–1676.

Perahia D, Mouawad L. Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Comput. Chem.* 1995, 19:241–246.

Pol-Fachin L, Rusu VH, Verli H, Lins RD. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* 2012, 8:4681-4690

Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SH, Presta LG. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J Biol Chem.* 2002, 277:26733–26740.

Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N, Shitara K. The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J Biol Chem.* 2003, 278:3466–3473.

Vanhooren V, Dewaele S, Kuro-o M, Taniguchi N, Dolle L, van Grunsven LA, Makrantonaki E, Zouboulis CC, Chen CC, Libert C. Alteration in N-glycomics during mouse aging: a role for FUT8. *Aging Cell* 2011, 10:1056–1066.

Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1996, 8:127-134.

Wang X, Fukuda T, Li W, Gao C, Kondo A, Matsumoto A, Miyoshi E, Taniguchi N, Gu J. Requirement of Fut8 for the expression of vascular endothelial growth factor receptor-2: a new mechanism for the emphysema-like changes observed in Fut8-deficient mice. *J. Biochem.* 2009, 145:643–651.

FIGURE CAPTION

Figure 1: RMSD analysis after 100ns MD for the coiled-coil (black), catalytic (red) and SH3 (blue) domains in FUT8 studied with the Amber force field without the substrate (free form, A), with the Gromos force-field also in its free form (B) and with the Gromos force field complexed to the substrate (C).

Figure 2: RMSF analysis (black) and Principal Component Analysis (vector 1, red, and vector 2, blue) after 100ns MD of FUT8 in its free form in the Amber force field (A), in the Gromos force field (B) and for FUT8 in its complexed form in the Gromos force field (C).

Figure 3. (A) RMSD analysis for the substrate GFB within the enzyme FUT8 in the absence of the five amino acid loop (black) and in the presence of the loop (red). (B) Initial (C atoms in green) and final (C atoms in cyan) conformations and orientations of GFB within FUT8 in the absence of the five amino acid loop. (C) Initial (C atoms in green) and final (C atoms in cyan) orientations of GFB within FUT8 in the presence of the five amino acid loop. We removed protein atoms in B and C for a better visualization.

Figure 4. Result obtained after molecular docking of the N-glycan (C atoms represented as cyan sticks) and FUT8-GDP- β -L-fucose (FUT8 is represented as blue surface, GDP- β -L-fucose C atoms are represented as green sticks).

Figure 5. Result obtained by molecular docking of FUT8 and Integrin $\alpha 5\beta 1$. (A) FUT8 (red) and chains A (green), B (cyan), C (pink/purple), D (yellow), E (violet/brown), F (grey), H (blue) and L (orange) of Integrin $\alpha 5\beta 1$. (B) Closer image of the interaction between FUT8 (red) and chain A (green) of Integrin $\alpha 5\beta 1$. Residue Asn275 in chain A, which is known to contain an N-glycan, is shown as spheres. Residues within FUT8 catalytic site (Asp353, His363, Ser469, and residues 484 to 491), which are linked to the enzyme substrate, GFB- β -L-fucose, are shown as sticks.

Figure 1

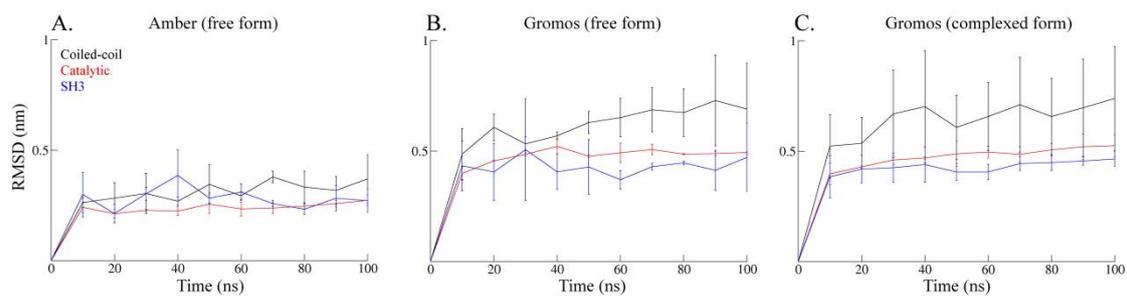


Figure 2

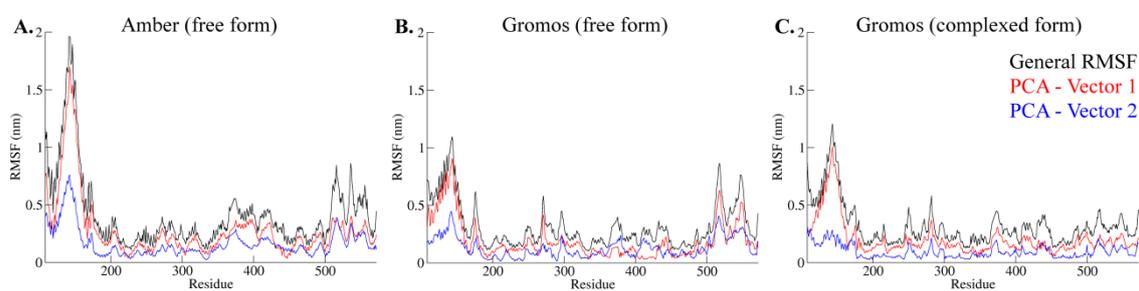


Figure 3

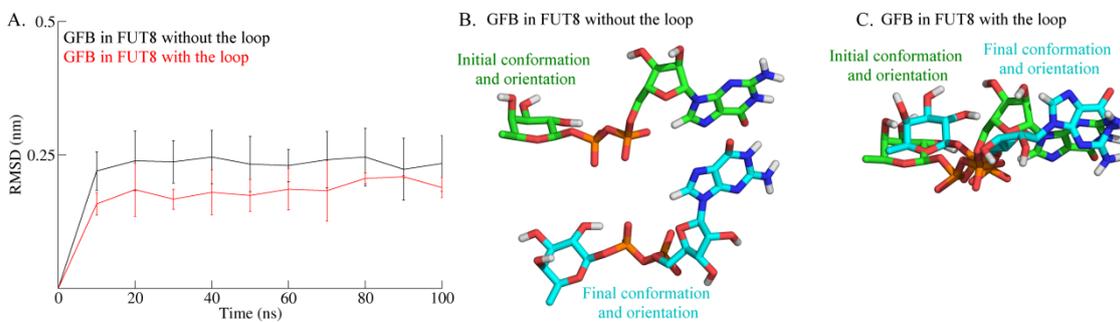


Figure 4.

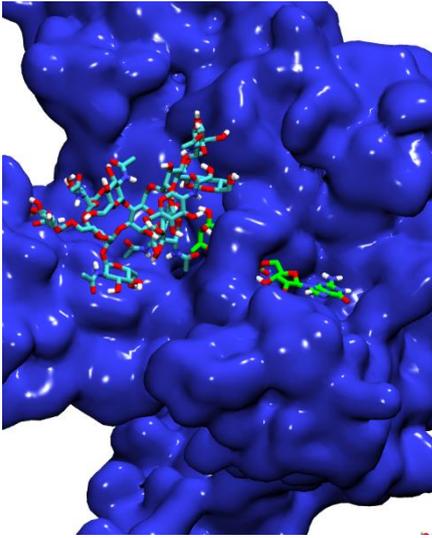
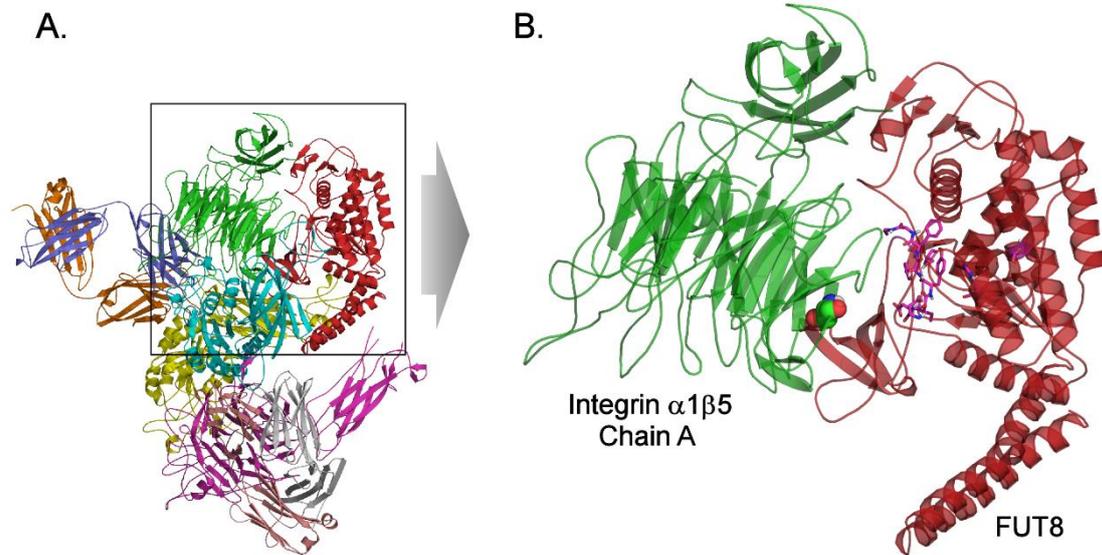


Figure 5.



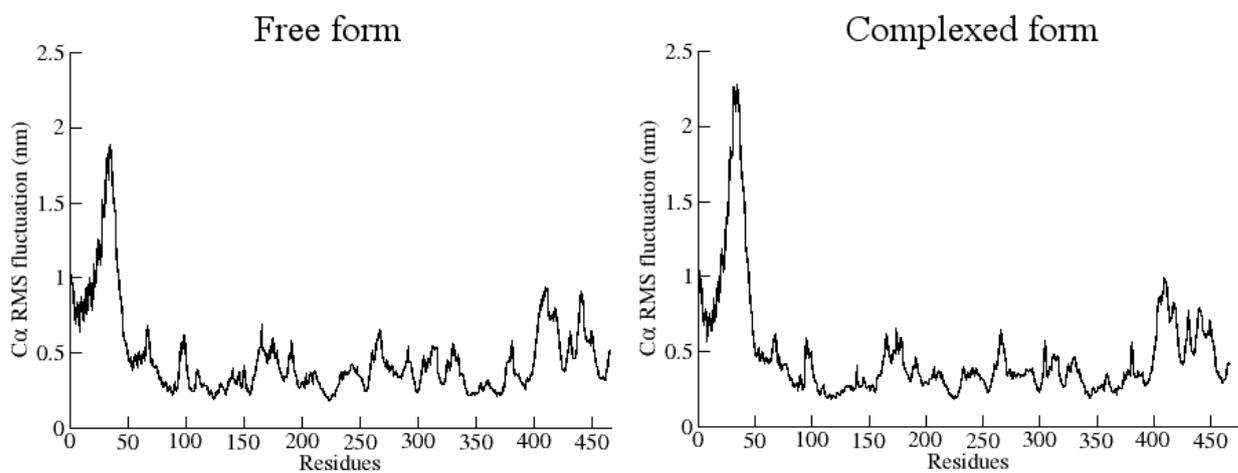
Supplementary data

Figure S1. RMSF after NMA for the free form and the complexed form of FUT8, confirming that the N-terminal part of FUT8 (coiled-coil) is a very flexible domain.

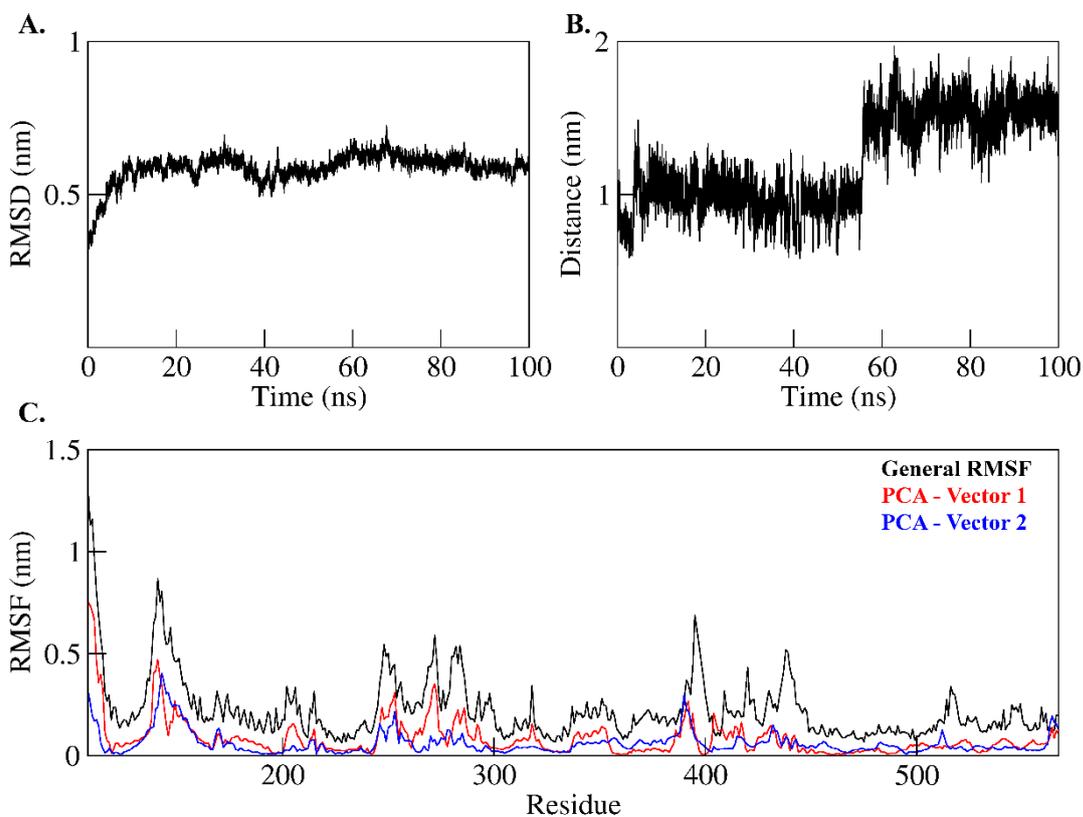


Figure S2. Analysis after MD of the complex FUT8-GFB- β -L-fucose and the N-glycan. **A.** RMSD after 100ns MD. **B.** Distance between GDP- β -L-fucose and the first GlcNAc from the N-glycan. **C.** RMSF and PCAs (vector 1 and vector 2) after 100ns MD.

4.4 Caracterização da glicobiologia estrutura de C2GnT-L.

4.2.1 DM de C2GnT-L livre e complexada aos seus substratos doador e acceptor

Após as simulações por DM (100ns) de C2GnT-L livre (PDBIDs 2gak e 2gam) e complexada a seus substratos acceptor (Gal- β -(1-3)-GalNAc, PDB 2gam) e doador (UDP-GlcNAc) foram realizadas diversas análises. A análise do desvio médio quadrático (RMSD) é apresentada na [Figura 17](#) para os quatro sistemas mencionados.

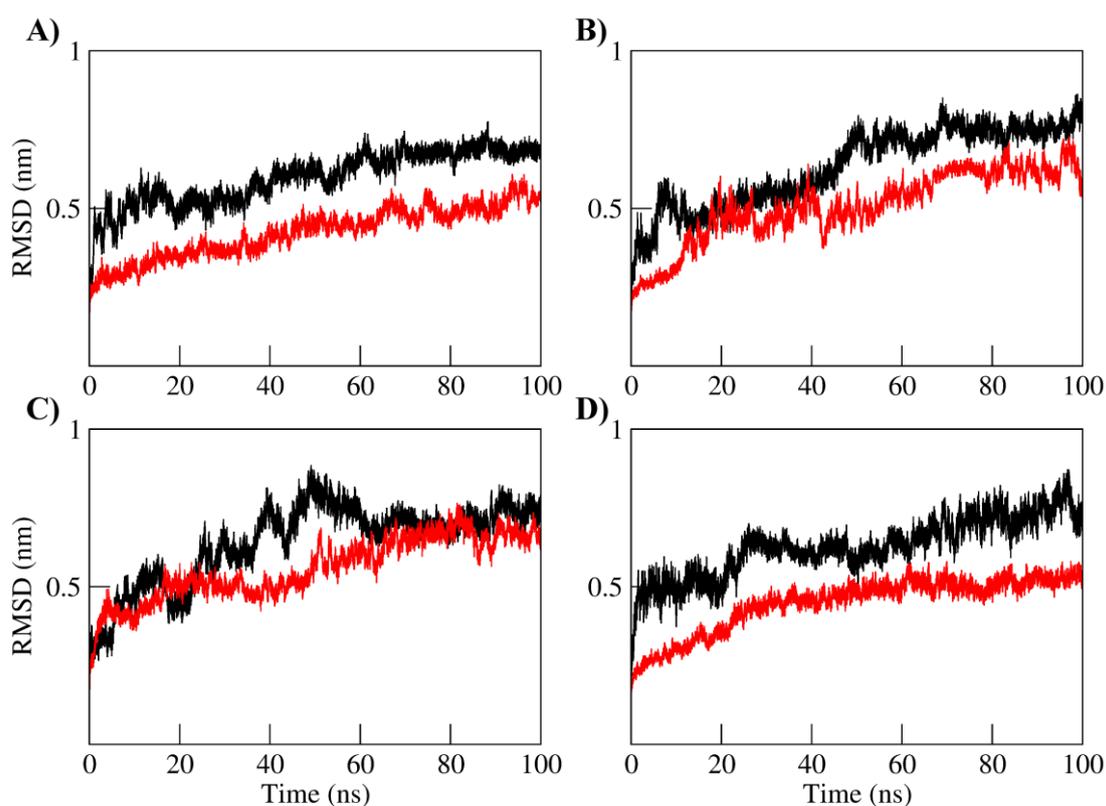


Figura 17. RMSD para C2GnT-L livre utilizando a estrutura de PDB 2gak (A), para C2GnT-L livre utilizando a estrutura de PDB 2gam (B), para C2GnT-L complexada a seu substrato acceptor, Gal- β -(1-3)-GalNAc, obtida do PDB 2gam (C), e para C2GnT-L complexada ao seu substrato doador, UDP-GlcNAc (D). As cores preta e vermelha correspondem respectivamente ao domínios N-terminal e catalítico.

Pode-se observar que em todos os sistemas o comportamento é semelhante. O domínio N-terminal, por ser composto de apenas α -hélices, apresenta um maior desvio (preto), enquanto o domínio catalítico permanece com um desvio mais baixo, não sofrendo alterações significativas com a ausência ou presença de ambos os substratos.

Foi também realizada a análise da flutuação média quadrática (RMSF) e a análise de componentes principais (PCA), apresentadas na [Figura 18](#) para os quatro sistemas.

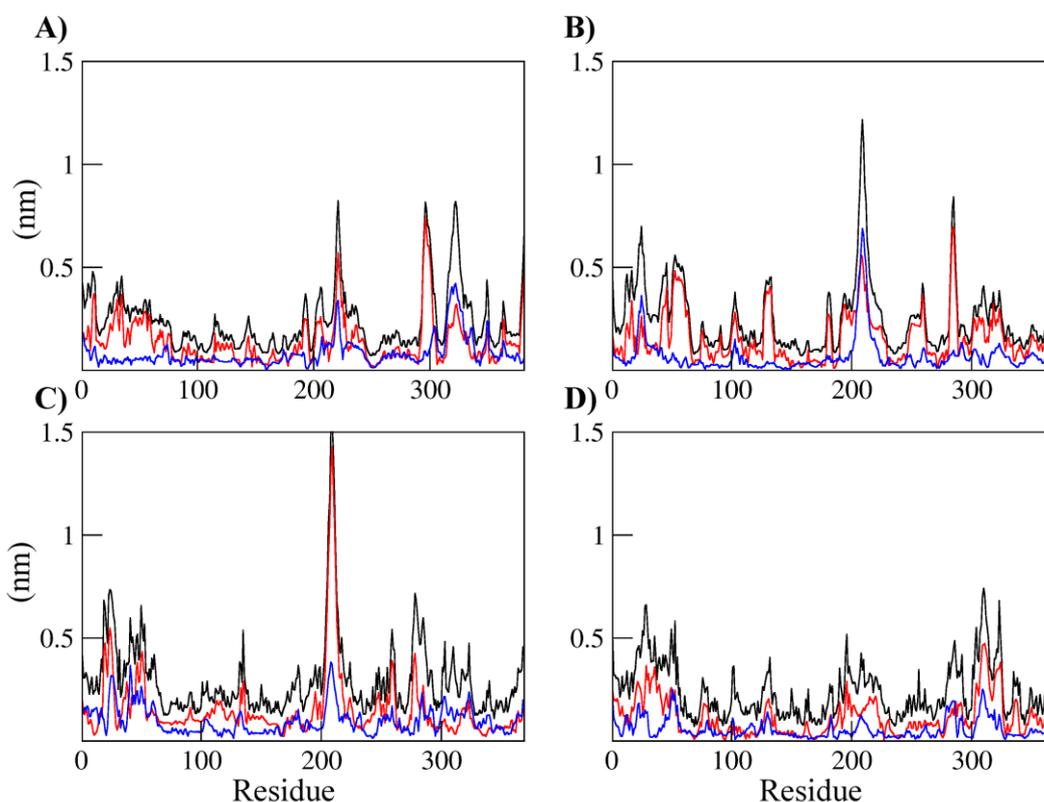


Figura 18. RMSF geral (preto) e análise de componentes principais (vetor 1 vermelho e vetor 2 azul) para C2GnT-L livre utilizando a estrutura de PDB 2gak (A), para C2GnT-L livre utilizando a estrutura de PDB 2gam (B), para C2gnT-L complexada a seu substrato aceptor, Gal- β -(1-3)-GalNAc, obtida do PDB 2gam (C), e para C2GnT-L complexada ao seu substrato doador, UDP-GlcNAc (D).

A partir destas análises, pode-se observar a flutuação média de cada resíduo. Cabe ressaltar aqui que no sistema apresentado em A, o único proveniente do PDB ID 2gak, há um deslocamento na numeração dos resíduos, correspondente a 12 resíduos, pois esta estrutura cristalográfica inicia-se no resíduo 44, enquanto nos demais, provenientes da estrutura 2gam, o início corresponde ao resíduo 56. Nesta análise observa-se que, em B e C, há uma região em torno do aminoácido 215-220 (que corresponde em A à região 227-232) com alta flexibilidade. Já em D esta região apresenta-se menos flexível.

Foi também realizada a análise da variação de estrutura secundária ao longo do tempo, apresentadas na [Figura 19](#) para os quatro sistemas.

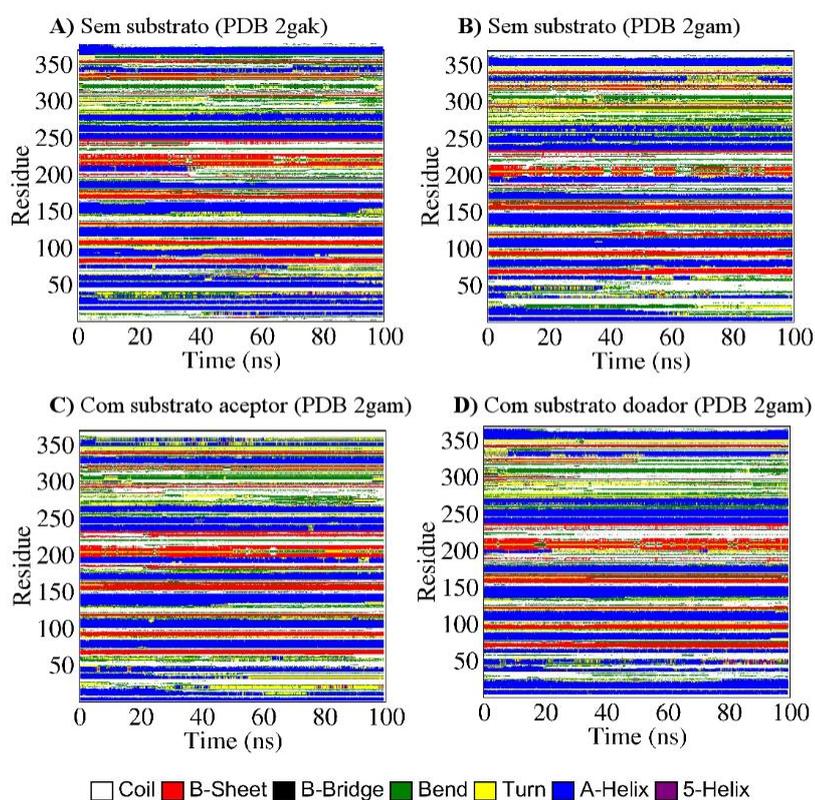
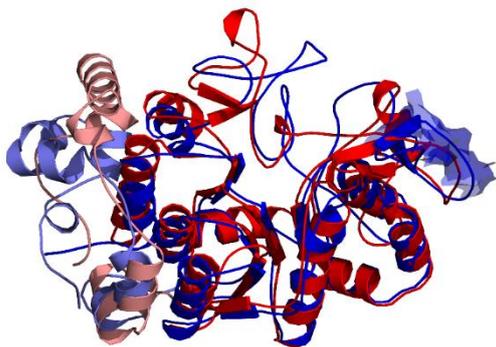


Figura 19. Variação da estrutura secundária durante o tempo de simulação por resíduo (conforme legenda de cores).

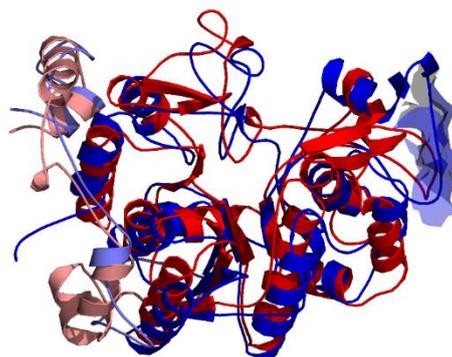
A partir destas análises, pode-se observar a variação de estrutura secundária por resíduo, ao longo do tempo de simulação. Relembrando, no sistema apresentado em A existem 12 resíduos adicionais no domínio N-terminal, o que desloca a numeração em 12.

A partir destas duas análises de flutuação, apresentadas nas figuras 13 e 14, procuramos a região de maior flexibilidade (resíduos 215 a 220) na visualização das estruturas ([Figura 20](#)).

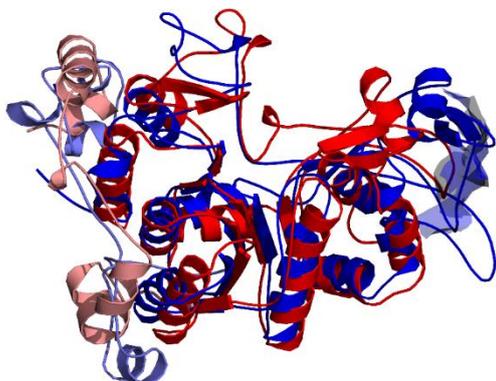
A) 2gak (sem substrato)



B) 2gam (sem substrato)



C) 2gam (com substrato aceptor)



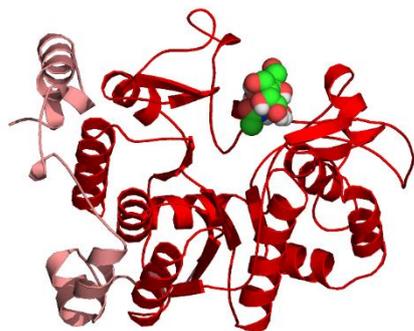
D) 2 gam (com substrato doador)



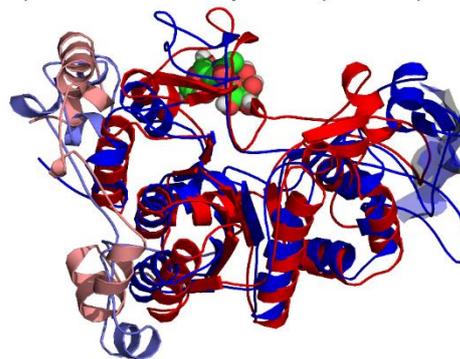
Figura 20. Visualização das estruturas iniciais (0ns, vermelho) e finais (100ns, azuis) de cada sistema. A região N-terminal está demonstrada em cores claras (salmão para 0ns e azul claro para 100ns). A região de maior flexibilidade (resíduos 215 a 220) está destacada pela superfície azul nas estruturas finais.

Pode-se observar que a região de maior flexibilidade (resíduos 215 a 220) corresponde a uma alça, portanto é esperado que haja uma alta flutuação. E, como observado nas figuras 19 e 20, não há perda de estrutura secundária nesta região. O único sistema que não apresentou alta flexibilidade nesta região foi o D. Isto provavelmente se deve ao posicionamento do substrato doador, como vemos na Figura 21 a seguir.

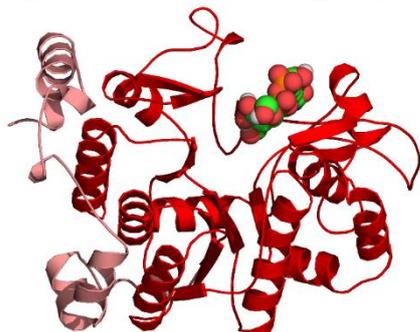
A) Substrato aceptor (0ns)



B) Substrato aceptor (100ns)



C) Substrato doador (0ns)



D) Substrato doador (100ns)

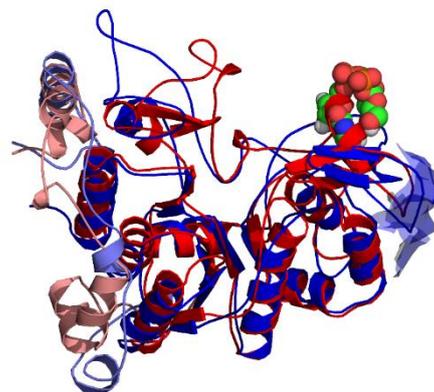


Figura 21. Visualização da estrutura inicial (0ns, vermelha) e final (100ns, azul) dos sistemas com substrato. Posição inicial (A) e final (B) do substrato aceptor (esferas) e posição inicial (C) e final (D) do substrato doador (esferas).

A intenção da [Figura 21](#) é demonstrar a posição de cada um dos substratos antes e depois da DM. Em [21A](#), observa-se a posição inicial do substrato aceptor, conforme ocorre na estrutura cristalográfica 2gam. Após a DM, foi observado que este substrato se desloca ([Figura 21B](#)). Uma análise na estrutura deste dissacarídeo mostra também que sua conformação se altera, indo de ângulos Φ e Ψ de -94° e -155° respectivamente (conformação cristalográfica, [Figura 22A](#)), para ângulos Φ e Ψ de respectivamente 60° e -120° ([Figura 22C](#)). Esta conformação final corresponde à conformação prevista em solução para este dissacarídeo ([Figura 22B](#)). Em [21C](#), temos a estrutura com o substrato doador obtida por atracamento molecular. Pode-se observar que este substrato ocupa inicialmente uma região similar à do substrato aceptor. Entretanto, após a DM, ocorre um deslocamento ([Figura 21D](#)) menor e para o lado contrário ao observado para o substrato aceptor. Utilizando a ferramenta

Ligplot do servidor PDBsum, obtivemos um detalhe acerca da ligação de cada um destes substratos em 0ns e 100ns (Tabela 1).

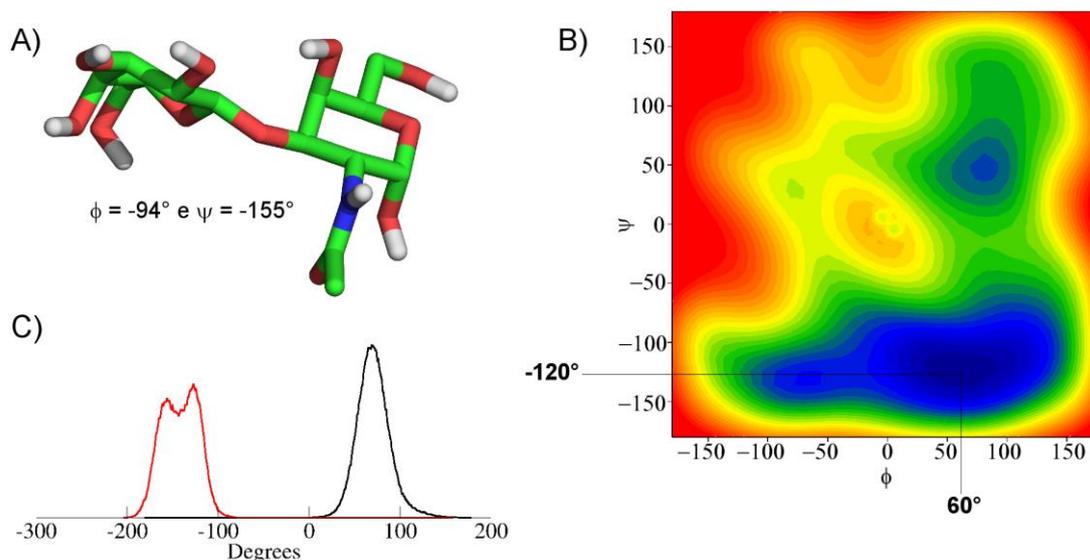


Figura 22. A) Conformação cristalográfica do substrato aceptor (Gal-β-(1-3)-GalNAc). B) Mapa de energia para este dissacarídeo após metadinâmica (regiões azuis = menor energia; regiões vermelhas = maior energia). C) Gráfico com os picos de prevalências conformacionais para o dissacarídeo após a DM em solução.

Tabela 1. Resíduos que interagem com os substratos

Estrutura cristalográfica (2gam)	Substrato aceptor (0ns) após minimização	Substrato aceptor (100ns)	Substrato doador (0ns)	Substrato doador (100ns)
			Tyr187*	
Glu243	Glu243		Glu243	
			Glu245*	
				Lys246*
			Pro248*	Pro248*
				Pro249*
Asn250			Asn250	Asn250
Lys251	Lys251		Lys251	Lys251
				Glu252*
Arg254			Arg254	Arg254
Gly285				
Ser286	Ser286			
				Tyr316*
Glu320	Glu320		Glu320	
			Ser345*	
Asp346	Asp346		Asp346	
Trp356	Trp356		Trp356	
Tyr358	Tyr358		Tyr358	
Phe359	Phe359	Phe359		
			Arg378*	
	Ser379*			
			Val380*	
		Gln394*		
		His395*		

*Resíduos não previstos na estrutura cristalográfica

A tabela 1 mostra que o substrato doador ocupa grande parte do sítio do substrato acceptor detectado pela cristalografia. Como o cristal não possui o substrato doador, é possível que, na ausência deste, o substrato acceptor tenha ocupado um espaço que não esteja realmente disponível, já que no momento da reação o substrato doador estará ali. Ainda, quando tentamos, por sobreposição, substituir o substrato acceptor cristalográfico, com os ângulos $\phi = -94^\circ$ e $\psi = -155^\circ$, pelo dissacarídeo com os ângulos obtidos por metadinâmica, não havia espaço para encaixe. Assim, é possível que este seja o sítio do substrato doador apenas.

5. DISCUSSÃO GERAL

5.1 Preâmbulo

Inicialmente, continuando uma linha de pesquisa já solidificada em nosso laboratório, envolvendo planejamento racional de novos candidatos a protótipos de fármacos anticoagulantes, meu projeto de doutorado, iniciado em outubro de 2010, se intitulava “*Caracterização estrutural e conformacional dos complexos tenase e protrombinase da cascata de coagulação sanguínea*”. Em outubro de 2011, com exatamente 1 ano de projeto em andamento, um artigo de um grupo americano intitulado “*A proposed ternary complex model of prothrombinase with prothrombin: protein–protein docking and molecular dynamics simulations*” (LEE ET AL., 2011) foi publicado, englobando 50% do meu projeto inicial. Desta maneira, tive que tomar a decisão entre arriscar realizar os 50% restante do trabalho ou escolher um novo assunto.

Após um período de pesquisas, tendo já trabalhando no meu mestrado com glicoproteínas de importância patológica e terapêutica (as PGHSs 1 e 2, envolvidas no processo inflamatório) e dada a minha formação em farmácia, escolhi trabalhar novamente com proteínas glicosiladas, ou envolvidas no processo de glicosilação, mas desta vez focando em outra patologia: câncer. Pesquisando as GTs superexpressas em tumores, escolhi trabalhar com duas que já possuíam estrutura cristalográfica determinada: FUT8 e C2GnT-L.

Paralelamente, surgiu a oportunidade de atuar no projeto acerca do principal alérgeno do gato, Fel d 1, que também é uma proteína glicosilada, junto com o colega de laboratório Rodrigo Braun. Esse trabalho inicialmente seria secundário para ambos, mas nosso interesse acabou crescendo acerca do estudo desta proteína, se tornando tão significativo quanto a minha ideia de projeto com as GTs.

5.2 Glicobiologia estrutural da Fel d 1

A literatura acerca da importância estrutural da glicosilação na proteína Fel d 1 era escassa e controversa. Enquanto parte da literatura indicava que sua deglicosilação acarretava na perda de 60% da estrutura terciária, em estudos com

dicroísmo circular (VAILES ET AL., 1994), outros trabalhos, ao compararem a produção *in vitro* de IgE da Fel d 1 glicosilada e deglicosilada, não notavam diferenças significativas (SEPPÄLÄ ET AL., 2005). Estes trabalhos afirmavam que a antigenicidade a IgE era conformacionalmente dependente (VAILES ET AL., 1994; KROLL-KRISTENSEN ET AL., 1997; SEPPÄLÄ ET AL., 2005), o que nos instigou a verificar se a glicosilação influenciava de alguma forma a dinâmica desta proteína.

Ao iniciar o estudo da Fel d 1 a partir da estrutura cristalográfica disponível (PDBID 2ejn), nos deparamos com os putativos sítios de ligação a íons Ca^{2+} , deduzidos pelos autores a partir de similaridade estrutural com a uteroglobina. Entretanto, nesta proteína não existe um consenso na literatura em relação a ligação à Ca^{2+} (ANDERSSON ET AL., 1994; NORD ET AL., 1995; BARNES ET AL., 1996; CHOWDHURY ET AL., 2002; KAISER ET AL., 2007). Analisando a metodologia utilizada para a cristalografia da Fel d 1, percebe-se que há uma concentração de íons Ca^{2+} muito maior (25 vezes) do que a encontrada no gato (ERJAVEC E STANOVNIK, 1987). Além disso, este dímero de Fel d 1 (ou tetrâmero, já que é um dímero de heterodímeros) foi construído de maneira recombinante, o que poderia causar alguma alteração, embora tenha demonstrado propriedades idênticas à proteína natural (KAISER ET AL., 2007).

No estudo de Fel d 1 e sua interação com íons Ca^{2+} , tanto por cálculos semiempíricos quanto por DM em três diferentes campos de força, foi possível constatar que o único cálcio que permanece ligado à proteína é o cálcio central. Os dois outros (laterais) ficam livres e saem dos sítios previstos durante as simulações. Tais resultados apontam fortemente para os íons Ca^{2+} destes sítios laterais como artefatos da estrutura cristalina.

Para o estudo de Fel d 1 glicosilada, foram comparadas duas formas diferentes de glicosilação. Uma delas contendo a glicana completa [(-Asn-)GlcNAc₂Man₃GlcNAc₃Gal₃(Fuc1)-NeuAc₃] e outra contendo a forma “truncada” da glicana, [(-Asn-)GlcNAc₂Man₃GlcNAc₃Gal₂], que corresponde a forma mínima descrita por espectrometria de massas (KROLL-KRISTENSEN ET AL., 1997). A forma glicosilada completa apresentou resultados gerais semelhantes à proteína não-glicosilada, ocorrendo alteração apenas no tamanho da cavidade de uma das subunidades da Fel d 1 (forma completamente glicosilada diminuiu uma das cavidades, conforme tabela IV do item 4.2). Já a forma com a glicana reduzida apresentou o desenovelamento parcial de duas hélices em uma das subunidades da

Fel d 1, o que acarretou um aumento de flexibilidade. Isso poderia indicar regiões propensas à perda de estrutura, como proposto por Vailes et al., em 1994, já que a diminuição no tamanho da glicana acarretou um desenovelamento de duas hélices. Mas, ainda assim, isso não foi observado na nossa amostragem da proteína não-glicosilada.

5.3 Glicobiologia estrutural de FUT8

A GT FUT8 está envolvida em diversos processos patológicos, como câncer, envelhecimento, e até mesmo fenótipos como enfisema e esquizofrenia (MIYOSHI ET AL., 1997; WANG ET AL., 2009; FUKUDA ET AL., 2011; VANHOOREM ET AL., 2011). Esta fucosiltransferase não se assemelha às outras FUTs em estrutura nem em função (a fucosilação terminal, sintetizada pelas FUTs 1 a 7 e FUT9 está associada à metástase) (LISTINSKY ET AL., 2011). Entretanto, devido à alta conservação de três pequenas regiões da FUT8 com a FUT2, que participam da ligação ao substrato (a principal delas, após estudos de mutagênese sítio dirigida, a região Pro358-Arg365), foi possível encaixar o substrato da enzima disponível para FUT2 (PDBID 3zy5) na estrutura da FUT8 por sobreposição (IHARA ET AL., 2007).

O estudo da FUT8 nas suas formas livre e complexada ao substrato demonstrou uma alta flexibilidade no domínio N-terminal (“*coiled-coil*”) em todos os sistemas, nos três campos de força (Amber e Gromos na DM e Charmm nos modos normais), evidenciando que este resultado não depende de campo de força ou do método de amostragem conformacional. Houve perda de hélices nesta região apenas no campo de força Gromos, e não no Amber, o que está de acordo com o que temos observado no nosso grupo de pesquisas e na literatura, onde as estruturas simuladas no Gromos, por ser um campo átomo unido, alcançam conformações na amostragem mais rapidamente do que as simuladas no Amber, que é um campo “*all-atom*” (HSIEH E LUO, 2010). Cabe ressaltar aqui que, embora a estrutura cristalográfica não contenha a região transmembrana desta enzima (pois inicia no resíduo 108), ela corresponde apenas aos resíduos 1 a 29, sendo que os demais resíduos ausentes, 30 a 107, correspondem a pequenas hélices e voltas. Desta maneira, acredita-se que a alta flexibilidade N-terminal ocorra mesmo na presença desta região não cristalizada.

Neste trabalho também verificamos a importância de uma região de 5 aminoácidos (alça DKVGT, resíduos 368-372) ausente na estrutura cristalográfica. Quando esta estrutura está ausente, temos um deslocamento do substrato no sítio catalítico, enquanto sua presença estabiliza a interação substrato-proteína. A presença de alças em GTs auxiliando a catálise é descrita na literatura (IHARA ET AL., 2007) e este pode ser o caso da GT FUT8.

Foram também realizados estudos de atracamento molecular da FUT8 a sua glicana alvo da fucosilação e a uma de suas proteínas alvo, a integrina $\alpha 5\beta 1$. Foi possível prever a região de interação da FUT8 com estas duas moléculas. A integrina alvo foi atracada à FUT8 na forma não glicosilada devido a limitações do servidor utilizado. Entretanto, é desejável, posteriormente, tentar encaixar a glicana alvo nesta estrutura, a fim de refinar o modelo deste complexo.

5.4 Glicobiologia estrutura de C2GnT-L

A enzima C2GnT-L é uma GT cuja superexpressão está correlacionada com a progressão tumoral de vários tipos de câncer (pulmão, próstata e colo-retal), tornando esta enzima extremamente interessante no estudo de metástase tumoral (MACHIDA ET AL., 2001; HAGISAWA ET AL., 2005). Ela apresenta dois sítios de N-glicosilação em sua região N-terminal (Asn58 e Asn95) sendo a glicosilação em Asn95 essencial para a atividade biológica de C2GnT-L - gerar ligantes funcionais de selectinas (HAMON ET AL., 2005).

A caracterização de C2GnT-L em suas formas livre e complexada a seus substratos doador (UDG-GlcNAc) ou acceptor (Gal- β -(1-3)-GlcNAc), por DM, demonstrou que o comportamento geral da enzima não se altera com a presença ou ausência dos substratos. Entretanto, houveram diferenças na flexibilidade da enzima quando complexada ao substrato doador ou acceptor. Foi possível verificar, após experimento de atracamento molecular, que o sítio preferencial para o substrato doador correspondia, em grande parte, ao mesmo sítio descrito para o substrato acceptor, por cristalografia de raios-X (PDBID 2gam). Considerando que o cristal em questão foi obtido na ausência do substrato doador e considerando também sua resolução, de 2,7Å, que pode ser considerada baixa, é possível que este sítio pertença, na realidade, ao substrato doador. Isso estaria de acordo com o resultado

obtido por DM, onde o substrato acceptor se desloca bastante em relação à sua posição inicial, enquanto o substrato doador, embora também se desloque um pouco, permanece interagindo com vários resíduos de sua conformação inicial. O posicionamento final do substrato doado, apenas levemente deslocado, e para uma direção contrária ao deslocamento do substrato acceptor, explica a diminuição da flexibilidade na região dos aminoácidos 215 a 220.

Em relação a este trabalho, pretende-se ainda realizar a caracterização da enzima glicosilada, já que a glicosilação em Asn95 é essencial para gerar ligantes de selectinas.

5.5 Impacto de múltiplas simulações por DM

Uma das dúvidas constantes quando se trabalha com DM é saber quando finalizar a simulação. Por quanto tempo deve-se simular para ter uma amostragem ideal? Devo estender minha simulação? Um artigo de 2010, de Lange e colaboradores, realizou um estudo comparativo entre diferentes campos de força e demonstrou que, muitas vezes, em simulações por longos tempos, como centenas de nanossegundos, corre-se o risco de a molécula estudada atingir uma conformação estável inativa ou de alta energia livre e dela permanecer neste estado por um longo tempo, o que distorceria a frequência populacional das conformações desta molécula (Lange et al., 2010). Uma estratégia para evitar isso seria, ao invés de estender o tempo de simulação, realizar múltiplas simulações simultaneamente.

Simulações replicadas geralmente não são idênticas, pois percorrem caminhos conformacionais diferentes. Entretanto, espera-se que em algum ponto ocorra a convergência. Réplicas independentes que resultam na observação de um mesmo fenômeno permitem definir a probabilidade deste fenômeno ocorrer. Segundo o estudo de Lange e colaboradores, as conformações previstas por RMN foram alcançadas mais rapidamente em várias simulações de 50ns, do que em uma simulação única de 1 μ s (Lange et al, 2010). Desta maneira, atingir-se-ia uma maior amostragem realizando várias simulações curtas, em vez de apenas uma longa simulação.

Esta foi a estratégia utilizada no trabalho com a enzima FUT8, cujo resultado encontra-se no item 4.3. Lá estes resultados são apresentados como médias e

desvios das simulações. Para exemplificar, vejamos os dados de RMSD para cada réplica do sistema FUT8-GDP-fucose (Figura 23).

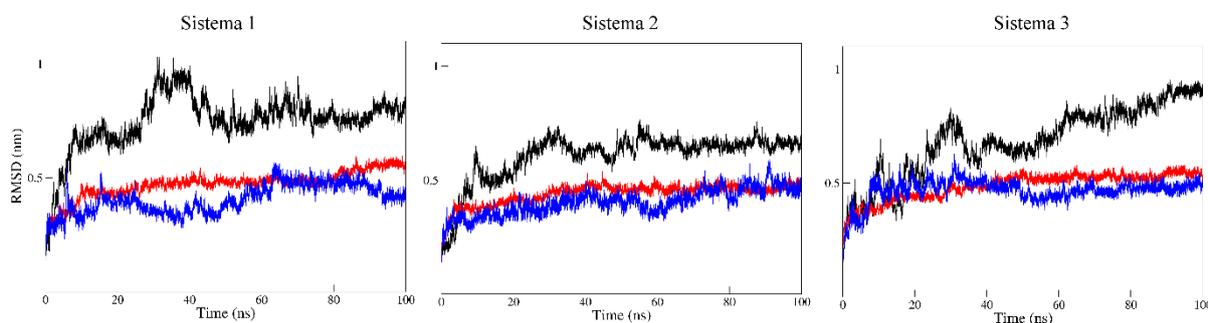


Figura 23. RMSD para os domínios N-terminal (*coiled-coil*, preto), catalítico (vermelho) e SH3 (azul) dos três sistemas de FUT8 complexada ao substrato GDP- β -L-fucose.

Percebe-se na Figura 23 que, embora os gráficos não sejam idênticos, o comportamento geral dos sistemas é muito semelhante, com a maior alteração em relação à estrutura de partida correspondendo ao domínio N-terminal (“*coiled-coil*”). A realização de réplicas, portanto, permite uma maior amostragem, dando uma noção da probabilidade dos fenômenos observados.

6. CONCLUSÕES

Considerando os objetivos propostos, o presente trabalho permitiu as seguintes conclusões:

- Foi possível utilizar ferramentas de química teórica para o estudo da proteína Fel d 1 em sua interação com íons Ca^{2+} e glicosilação, demonstrando por diferentes metodologias que os sítios laterais de Ca^{2+} parecem ser artefatos da cristalização da molécula, enquanto o íon Ca^{2+} central parece estar ligado à proteína. Ainda, pode-se observar que o comportamento geral da forma completamente glicosilada de Fel d 1 não diferiu da forma não-glicosilada, apresentando apenas uma redução no tamanho da cavidade de uma das subunidades. Já a proteína com glicosilação reduzida apresentou um aumento de flexibilidade em duas hélices em uma das subunidades da Fel d 1, devido a desenovelamento das mesmas.
- O estudo de FUT8 em solução por diferentes metodologias e em diferentes campos de força demonstrou que o movimento N-terminal é o principal componente da flexibilidade desta enzima, não sendo, portanto, dependente do campo de força ou do método de amostragem. A presença da alça de 5 aminoácidos DKVGT em FUT8 estabilizou a orientação do substrato GDP- β -L-fucose no sítio catalítico e, desta maneira, pode ser importante para o mecanismo da catálise. O atracamento molecular à glicana alvo da fucosilação e à integrina $\alpha 5\beta 1$ resultou em complexos compatíveis com a reação catalítica, além de fornecer informações sobre o sítio de ligação/interação desta enzima.
- A caracterização de enzima C2GnT-L demonstrou que o seu substrato doador se liga no mesmo sítio previsto por cristalografia de raios-X para o substrato aceptor, sendo que o substrato doador permanece mais estável no sítio durante a DM do que o aceptor, evidenciando que é possível que o sítio seja, na verdade, do substrato doador.

Além disso, o trabalho permitiu aplicar e consolidar diferentes metodologias computacionais no estudo destas diferentes proteínas, auxiliando na compreensão das relações estrutura-função das mesmas, contribuindo para sua aplicação biotecnológica em posteriores estudos de inibidores farmacológicos.

7. PERSPECTIVAS

Considerando-se os resultados obtidos no presente trabalho, as seguintes perspectivas podem ser traçadas:

- ▶ Analisar a proteína Fel d 1 da perspectiva do gato, considerando seu papel na fisiologia e no comportamento felino, que são aspectos largamente ignorados na literatura especializada;
- ▶ Estudar a enzima FUT8 complexada à integrina $\alpha 5\beta 1$ por DM;
- ▶ Caracterizar a enzima C2GnT-L em sua forma glicosilada (em Asn58 e Asn95), a fim de verificar seu papel na dinâmica desta GT.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Abbas, A. K.; Lichtman, A. H.; and Pillai, S. *Cellular and molecular immunology*, 7th ed, **2012**.
- Altmann, F. The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.*, 14:99-115, **2007**.
- Alon, R. and Rosen, S. Rolling on N-linked glycans: a new way to present L-selectin binding sites. *Nat. Immunol.*, 8:339–341, **2007**.
- Andersson, O.; Nordlund-Moller, L.; Barnes, H. J.; and Lund, J. Heterologous expression of human uteroglobin/polychlorinated biphenyl-binding protein. Determination of ligand binding parameters and mechanism of phospholipase A2 inhibition *in vitro*. *J. Biol. Chem.*, 269:19081–19087, **1994**.
- Barnes, H. J.; Nordlund-Moller, L.; Nord, M.; Gustafsson, J.; Lund, J.; and Gillner, M. Structural basis for calcium binding by uteroglobins. *J. Mol. Biol.*, 256:392–404, **1996**.
- Baycin-Hizal, D.; Gottschalk, A.; Jacobson, E.; Mai, S.; Wolozny, D.; Zhang, H.; Krag, S. S.; Betenbaugh, M. J. Physiologic and pathophysiologic consequences of altered sialylation and glycosylation on ion channel function. *Biochem. Biophys. Res. Commun.*, in press, **2014**.
- Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; and Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684-3690, **1984**.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; and Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, **2000**.
- Berman, H. M.; Henric, K.; Nakamura, H.; and Markley, J. Building meaningful models of glycoproteins. *Nat. Struct. Mol. Biol.*, 14:354-355, **2007**.
- Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F. Jr.; Brice, M. D.; Rogers, J. R.; Kennard, O.; Shimanouchi, T.; and Tasumi, M. The Protein Data Bank: a computerbased archival file for macromolecular structures. *J. Mol. Biol.*, 112:535–547, **1977**.
- Bierhuizen, M. F.; and Fukuda, M. Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen. *Proc. Natl. Acad. Sci. U. S. A.* 89:9326–9330, **1992**.
- Bourne, Y.; and Henrissat, B. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.*, 11: 593–600, **2001**.
- Breton, C.; Bettler, E.; Joziase, D. H.; Geremia, R. A.; and Imberty, A. Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *J. Biochem.*, 123:1000–1009, **1998**.

Breton, C.; and Imberty, A. Structure-function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.*, 9, 563–571, **1999**.

Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koca, J.; and Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology*, 16:29R–37R, **2006**.

Case, D. A.; Cheatham, T. E. 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; and Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.*, 26:1668-1688, **2005**.

Chowdhury, B.; Mantile-Selvaggi, G.; Miele, L.; Cordella-Miele, E.; Zhang, Z.; Mukherjee, A. B. Lys43 and Asp46 in alpha-helix 3 of uteroglobin are essential for its phospholipase A2 inhibitory activity. *Biochem. Biophys. Res. Commun.*, 295:877-883, **2002**.

Comeau, S. R.; Gatchell, D. W.; Vajda, S.; and Camacho, C. J. ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.*, 32:W96–W99, **2004**.

Coutinho, P. M.; Deleury, E.; Davies, G. J.; and Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, 328:307–317, **2003**.

Crispin, M.; Stuart, D. I.; and Jones, E. Y. Building meaningful models of glycoproteins. *Nat. Struct. Mol. Biol.*, 14:354, **2007**.

Cumming, D. A.; and Carver, J. P. Virtual and solution conformations of oligosaccharides. *Biochemistry*, 26:6664-6676, **1987**.

Darden, T.; York, D.; and Pedersen, L. Particle Mesh Ewald – an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089-10092, **1993**.

de Beer, T.; Vliegthart, J. F.; Löffler, A.; and Hofsteenge, J. The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase Us Is α -Mannopyranose. *Biochemistry*, 34:11785-11789, **1995**.

de Groot, B. L.; and Grubmüller, H.: Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science*, 294:2353-2357, **2001**.

DeLano, W. L.: The PyMOL Molecular Graphics System. DeLano Scientific, South San Francisco, CA, USA, **2002**.

de Sant'Anna, C. M. R.: Glossário de termos usados no planejamento de fármacos (recomendações IUPAC 1997). *Quim. Nova*, 25:505-512, **2002**.

Doucey, M. A.; Hess, D.; Cacan, R.; and Hofsteenge, J. Protein C-Mannosylation Is Enzyme-catalyzed and Uses Dolichyl-Phosphate-Mannose as Precursor. *Mol. Biol. Cell*, 9:291-300, **1998**.

Drickamer, K.; and Taylor, M. E. Biology of animal lectins. *Annu. Rev. Cell Biol.*, 9:237-264, **1993**.

Duffort, O. A.; Carreira, J.; Nitti, G.; Polo, F.; and Lombardero, M. Studies on the biochemical structure of the major cat allergen *Felis domesticus* I. *Mol. Immunol.*, 28:301–309, **1991**.

Dwek, R. A. Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev.*, 96:683-720, **1996**.

Emara, M.; Royer, P.-J.; Abbas, Z.; Sewell, H. F.; Mohamed, G. G.; Singh, S.; Peel, S.; Fox, J.; Shakib, F.; Martinez-Pomares, L.; and Ghaemmaghami, A. M. Recognition of the Major Cat Allergen Fel d 1 through the Cysteine-rich Domain of the Mannose Receptor Determines Its Allergenicity. *J. Biol. Chem.*, 286:13033-13040, **2011**.

Erjavec, F.; Stanovnik, L. Histamine and calcium secretion in the submandibular gland of the cat evoked by two types of stimulation. *Agents Actions*, 20:188-190, **1987**.

Feizi, T. Carbohydrate antigens in human cancer. *Cancer Surv.*, 4:245–269, **1985**.

Fernandes, C. L.; Sachett, L. G.; Pol-Fachin, L.; Verli, H. GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins. *Carbohydr Res.*, 345:663-671, **2010**.

Fiser, A.; Do, R.K.; and Sali, A. Modeling of loops in protein structures. *Protein Science*, 9:1753-1773, **2000**.

Fukuda, T.; Hashimoto, H.; Okayasu, N.; Kameyama, A.; Onogi, H.; Nakagawasai, O.; Nakazawa, T.; Kurosawa, T.; Hao, Y.; Isaji, T.; Tadano, T.; Narimatsu, H.; Taniguchi, N.; and Gu, J. 1,6-Fucosyltransferase-deficient Mice Exhibit Multiple Behavioral Abnormalities Associated with a Schizophrenia-like Phenotype. *J. Biol. Chem.*, 286:18434–18443, **2011**.

Fuster, M. M.; and Esko, J. D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nat. Rev. Cancer*, 5:526-542, **2005**.

Grönlund, H.; Saarne, T.; Gafvelin, G.; and van Hage, M. The Major Cat Allergen, Fel d 1, in Diagnosis and Therapy. *Int. Arch. Allergy Immunol.*, 151:265-274, **2010**.

Guex, N.; and Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714-2723, **1997**.

Hagisawa, S.; Ohyama, C.; Takahashi, T.; Endoh, M.; Moriya, T.; Nakayama, J.; Arai, Y.; and Fukuda, M. Expression of core 2 β 1,6-N-acetylglucosaminyltransferase facilitates prostate cancer progression. *Glycobiology* ,15:1016–1024, **2005**.

Hanson, S. R.; Culyba, E. K.; Hsu, T. L.; Wong, C. H.; Kelly, J. W.; Powers, E. T. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. U. S. A.*, 106:3131-3136, **2009**.

Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463-1472, **1997**.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435-447, **2008**.

Hofsteenge, J.; Müller, D. R.; de Beer, T.; Löffler, A.; Richter, W. J.; and Vliegthart, J. F. New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry*, 33:13524-13530, **1994**.

Holm, L.; and Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, 233:123-138, **1993**.

Hsieh, M-J.; and Luo, R. Balancing simulation accuracy and efficiency with the Amber united atom force field. *J. Phys. Chem. B*, 114:2886-2893, **2010**.

Humphrey, W.; Dalke, A.; and Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33-38, **1996**.

Ihara, H.; Ikeda, Y.; Toma, S.; Wang, X.; Suzuki, T.; J Gu, J.; Miyoshi, E.; Tsukihara, T.; Honke, K.; Matsumoto, A.; Nakagawa, A.; and Taniguchi, N. Crystal structure of mammalian α 1,6-fucosyltransferase, FUT8. *Glycobiology* 17:455–466, **2007**.

IUPAC. *Compendium of Chemical Terminology*, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: <http://goldbook.iupac.org> (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins.

Johnson, J. L.; Jones, M. B.; Ryan, S. O.; and Cobb, B. A. The regulatory power of glycans and their binding partners in immunity. *Trends Immunol.*, 34:290-298, **2013**.

Julenius, K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology*, 17:868-876, **2007**.

Kaiser, L.; Velickovic, T. C.; Badia-Martinez, D.; Adedoyin, J.; Thunberg, S.; Hallén, D.; Berndt, K.; Grönlund, H.; Gafvelin, G.; van Hage, M.; and Achour, A. Structural Characterization of the Tetrameric form of the Major Cat Allergen Fel d 1. *J. Mol. Biol.*, 370:714–727, **2007**.

Kawashima, H.; Petryniak, B.; Hiraoka, N.; Mitoma, J.; Huckaby, V.; Nakayama, J.; Uchimura, K.; Kadomatsu, K.; Muramatsu, T.; Lowe, J. B.; and Fukuda, M. N-acetylglucosamine-6-O-

sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial. *Nat. Immunol.*, 6:1096–1104, **2005**.

Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; and Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181:662–666, **1985**.

Klamt, A.; and Schüümann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans.*, 2: 799–805, **1993**.

Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*, 65:392–406, **2006**.

Krieg, J.; Gläsner, W.; Vicentini, A.; Doucey, M. A.; Löffler, A.; Hess, D.; and Hofsteenge, J. C-Mannosylation of Human RNase 2 Is an Intracellular Process Performed by a Variety of Cultured Cells. *J. Biol. Chem.*, 272:26687–26692, **1997**.

Krieg, J.; Hartmann, S.; Vicentini, A.; Gläsner, W.; Hess, D.; and Hofsteenge, J. Recognition Signal for C-Mannosylation of Trp-7 in RNase 2 Consists of Sequence Trp-x-x-Trp. *Mol. Biol. Cell*, 9:301–309, **1998**.

Kroll-Kristensen, A. K.; Schou, C.; Roepstorff, P. Determination of isoforms, N-linked glycan structure and disulfide bond linkages of the major cat allergen Fel d 1 by mass spectrometric approach. *Biol. Chem.*, 378:899–908, **1997**.

Lange, O. F.; van der Spoel, D.; and de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.*, 99:647–655, **2010**.

Leach, A. R.: *Molecular Modelling Principles and Applications*, 2nd Ed., Longman, Cingapura, **2001**.

Lee, C. J.; Wu, S.; and Pedersen, L. G. A proposed ternary complex model of prothrombinase with prothrombin: protein–protein docking and molecular dynamics simulations. *J. Thromb. Haemost.*, 9:2123–2126, **2011**.

Lee, S. H.; Takahashi, M.; Honke, K.; Miyoshi, E.; Osumi, D.; Sakiyama, H.; Ekuni, A.; Wang, X.; Inoue, S.; Gu, J. P.; Kadomatsu, K.; Taniguchi, N. Loss of core fucosylation of low-density lipoprotein receptor-related protein-1 impairs its function, leading to the upregulation of serum levels of insulin-like growth factor-binding protein 3 in Fut8^{-/-} mice. *J. Biochem.*, 139:391–398, **2006**.

Leitermann, K.; and Ohmann, J. L. Cat allergen 1: Biochemical, antigenic and allergic properties. *J. Allergy Clin. Immunol.*, 74:147–153, **1984**.

Li, W.; Nakagawa, T.; Koyama, N.; Wang, X.; Jin, J.; Mizuno-Horikawa, Y.; Gu, J.; Miyoshi, E.; Kato, I.; Honke, K.; Taniguchi, N.; Kondo, A. Down-regulation of trypsinogen expression is associated with

growth retardation in α 1,6-fucosyltransferase- deficient mice: attenuation of proteinase-activated receptor 2 activity. *Glycobiology*, 16:1007–1019, **2006**.

Ligabue-Braun, R. Biologia estrutural de ureases: filogenia, ativação e peptídeos derivados. Tese (Doutorado em Ciências) – Centro de Biotecnologia, UFRGS, Porto Alegre, **2014**.

Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78:1950–1958, **2010**.

Listinsky, J. J.; Siegal, G. P.; and Listinsky, C. M. The emerging importance of α -L-fucose in human breast cancer: a review. *Am. J. Transl. Res.*, 3:292-322, **2011**.

Litynska, A.; Przybylo, M.; Pochee, E.; and Laidler, P. Adhesion properties of human bladder cell lines with extracellular matrix components: the role of integrins and glycosylation. *Acta Biochim. Pol.*, 49:643–650, **2002**.

Low, M. G. Glycosyl-phosphatidylinositol: a versatile anchor for cell surface proteins. *FASEB J.*, 3:1600-1608, **1989**.

Lowe, J. B. Glycan-dependent leukocyte adhesion and recruitment in inflammation. *Curr. Opin. Cell Biol.*, 15:531–538, **2003**.

Lütteke, T.; Frank, M.; and von der Lieth, C. -W. Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydr. Res.*, 339:1015-1020, **2004**.

McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of folded proteins. *Nature*, 267:585-590, **1997**.

Machida, E.; Nakayama, J.; Amano, J.; and Fukuda, M. Clinicopathological Significance of Core 2 β 1,6-N-Acetylglucosaminyltransferase Messenger RNA Expressed in the Pulmonary Adenocarcinoma Determined by in Situ Hybridization. *Cancer Res.* 61:2226–2231, **2001**.

MackKerell, A. D. Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L. Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. 3rd; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102:3586-3616, **1998**.

MackKerell, Jr. A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56:257–265, **2001**.

Maggin, E. J.; Elliott, J. R.: Historical perspective and current outlook for Molecular Dynamics as a chemical engineering tool. *Ind. Eng. Chem. Res.*, 49:3059-3078, **2010**.

Maluf, F.V., Muniz, J.R.C., Oliva, G., Guido, R.V.C.: *Cristalografia de proteínas*. In Verli, H (Ed.) *Bioinformática: da biologia molecular à flexibilidade molecular*. Porto Alegre: e-book (<http://www.ufrgs.br/bioinfo/ebook/>), 2014.

Marth, J. Glycosylation changes in ontogeny and cell activation. in *Essentials of Glycobiology* (eds Varki, A. et al.) 515–536 (Cold Spring Harbor Laboratory Press, New York), **1999**.

Matsushita, Y.; Cleary, K. R.; Ota, D. M.; Hoff, S. D.; and Irimura, T. Sialyl-dimeric Lewis-X antigen expressed on mucin-like glycoproteins in colorectal-cancer metastases. *Lab. Invest.*, 63:780–791, **1990**.

Meezan, E.; Wu, H. C.; Black, P. H.; and Robbins, P. W. Comparative studies on carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry*, 8:2518–2524, **1969**.

Miyoshi, E.; Uozumi, N.; Noda, K.; Hayashi, N.; Hori, M.; and Taniguchi, N. Expression of α 1-6 Fucosyltransferase in rat tissues and human cancer cell lines. *Int. J. Cancer*, 72:1117–1121, **1997**.

Miyoshi, E.; Noda, K.; Ko, J. H.; Ekuni, A.; Kitada, T.; Uozumi, N.; Ikeda, Y.; Matsuura, N.; Sasaki, Y.; Hayashi, N.; Hori, M.; and Taniguchi, N. Overexpression of α 1–6 fucosyltransferase in hepatoma cells suppresses intrahepatic metastasis after splenic injection in athymic mice. *Cancer Res.*, 59:2237–2243, **1999**.

Monies, D. M.; Al-Hindi, H. N.; Al-Muhaizea, M. A.; Jaroudi, D. J.; Al-Younes, B.; Naim, E. A.; Wakil, S. M.; Meyer, B. F.; Bohlega, S. Clinical and pathological heterogeneity of a congenital disorder of glycosylation manifesting as a myasthenic/myopathic syndrome. *Neuromuscul. Disord.*, 24:353–359, **2014**.

Moran, A. P.; Gupta, A.; Joshi, L. Sweet-talk: Role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut*, 60:1412–1425, **2011**.

Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, 16:2785-2791, **2009**.

Nakagawa, H.; Zheng, M.; Hakomori, S.; Tsukamoto, Y.; Kawamura, Y.; and Takahashi, N. Detailed Oligosaccharide Structures of Human Integrin α 5 β 1 Analyzed by a Three-Dimensional Mapping Technique. *Eur. J. Biochem.*, 237:76–85, **1996**.

Noda, K.; Miyoshi, E.; Uozumi, N.; Yanagidani, S.; Ikeda, Y.; Gao, C.; Suzuki, K.; Yoshihara, H.; Yoshikawa, K.; Kawano, K.; Hayashi, N.; Hori, M.; Taniguchi, N. Gene expression of α 1–6 fucosyltransferase in human hepatoma tissues: a possible implication for increased fucosylation of α -fetoprotein. *Hepatology*, 28:944–952, **1998**; Erratum in: *Hepatology*, 29:301, **1999**.

Nord, M.; Gustafsson, J. A.; and Lund, J. Calcium dependent binding of uteroglobin (PCB-BP/CCSP) to negatively charged phospholipid liposomes. *FEBS Letters*, 374:403–406, **1995**.

Ohtsubo, K.; Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, 126:855–867, **2006**.

Okazaki, A.; Shoji-Hosaka, E.; Nakamura, K.; Wakitani, M.; Uchida, K.; Kakita, S.; Tsumoto, K.; Kumagai, I.; Shitara, K. Fucose depletion from human IgG1 oligosaccharide enhances binding enthalpy and association rate between IgG1 and FcγRIIIa. *J Mol Biol.*, 336:1239–1249, **2004**.

Oostenbrink, C.; Villa, A.; Mark, A. E.; and van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25:1656-1676, **2004**.

Pak, J. E.; Arnoux, P.; Zhou, S.; Sivarajah, P.; Satkunarajah, M.; Xing, X.; and Rini, J. M. X-ray Crystal Structure of Leukocyte Type Core 2 1,6-N-Acetylglucosaminyltransferase: Evidence for a Convergence of Metal Ion-Independent Glycosyltransferase Mechanism. *J. Biol. Chem.*, 281:26693–26701, **2006**.

Pancino, G.; Osinaga, E.; Charpin, C.; Mistro, D.; Barque, J. P.; and Roseto, A. Purification and characterization of a breast-cancer-associated glycoprotein not expressed in normal breast and identified by monoclonal antibody-83d4. *Brit. J. Cancer* 63:390–398, **1991**.

Parekh, R. B.; Dwek, R. A.; Thomas, J. R.; Rademacher, T. W.; Opdenakker, G.; Wittwer, A. J.; Howard, S. C.; Nelson, R.; Siegel, N. R.; Jennings, M. G.; Hatakas, N. K.; Feder, J. – Cell-type-specific and site-specific N-glycosylation of type I and type II human tissue plasminogen activator. *Biochemistry*, 28:7644-7662, **1989**.

Paulick, M. G.; Bertozzi, C. R. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. *Biochemistry*, 47:6991-7000, **2008**.

Pérez, S.; and Mulloy, B. Prospects for glycoinformatics. *Curr. Opin. Struc. Biol.*, 15:517-524, **2005**.

Petsko, G. A.; and Ringe, D. Protein Structure and Function. Oxford University Press, 1ed., **2009**.

Pol-Fachin, L.; and Verli, H. Assessment of glycoproteins dynamics from computer simulations. *Mini Rev. Org. Chem.*, 8:229-238, **2011**.

Pol-Fachin, L.; and Verli, H. Structural glycobiology of the major allergen of *Artemisia vulgaris* pollen, Art v 1: O-glycosylation influence on the protein dynamics and allergenicity. *Glycobiology*, 22:817-825, **2012**.

Pol-Fachin, L.; Rusu, V.H.; Verli, H.; Lins, R.D. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.*, 8:4681-4690, **2012**.

Ponder, J. W.; Case, D. A. Force fields for protein simulations. In *Advances in protein chemistry*, volume 66: protein simulations. Richards, F. M.; Eisenberg, D. S.; Kuriyan, J. Eds., Elsevier Academic Press, San Diego, pp 27-85, **2003**.

Price, N. J.; Pinheiro, C.; Soares, C. M.; Ashford, D. A.; Ricardo, C. P.; Jackson, P. A. A biochemical and molecular characterization of LEP1, an extensin peroxidase from lupin. *J. Biol. Chem.*, 278:41389-41399, **2003**.

Prorok-Hamon, M.; Notel, F.; Mathieu, S.; Langlet, C.; Fukuda, M.; and El-Battari, A. N-glycans of core2 $\beta(1,6)$ -N-acetylglucosaminyltransferase-I (C2GnT-I) but not those of $\alpha(1,3)$ -fucosyltransferase-VII (FucT-VII) are required for the synthesis of functional P-selectin glycoprotein ligand-1 (PSGL-1): effects on P-, L- and E-selectin binding. *Biochem. J.*, 391:491–502, **2005**.

Qasba, P. K. Involvement of sugars in protein-protein interactions. *Carbohydrate polymers*, 41:293-309, **2000**.

Robinson, L. N.; Artpradit, C.; Raman, R.; Shriver, Z. H.; Ruchirawat, M.; Sasisekharan, R. Harnessing glycomics technologies: Integrating structure with function for glycan characterization. *Electrophoresis*, 33:797–814, **2012**.

Sachett, L. G.; and Verli, H. Dynamics of different arachidonic acid orientations bound to prostaglandin endoperoxide synthases. *Eur. J. Med. Chem.*, 46:5212-5217, **2011**.

Saussez, S.; Marchant, H.; Nagy, N.; Decaestecker, C.; Hassid, S.; Jortay, A.; Schüring, M. P.; Gabius, H. J.; Danguy, A.; Salmon, I.; Kiss, R. Quantitative glycohistochemistry defines new prognostic markers for cancers of the oral cavity. *Cancer*, 82:252–260, **1998**.

Schou, C. Defining allergens of mammalian origin. *Clin. Esp. Allergy*, 23:7-14, **1993**.

Schuman, B.; Evans, S. V.; Fyles, T. M. Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *PLoS ONE* 8(8):e71077, **2003**.

Schwientek, T.; Yeh, J. C.; Levery, S. B.; Keck, B.; Merx, G.; van Kessel, A. G.; Fukuda, M.; and Clausen, H. Control of O-Glycan Branch Formation: molecular cloning and characterization of a novel thymus-associated core 2 $\beta(1,6)$ -n-acetylglucosaminyltransferase. *J. Biol. Chem.*, 275:11106–11113, **2000**.

Seppälä, U.; Häggglund, P.; Wurtzen, P. A.; Ipsen, H.; Thorsted, P.; Lenhard, T.; Roepstorff, P.; and Spangfort, M. D. Molecular Characterization of Major Cat Allergen Fel d 1. Expression of heterodimer by use of a baculovirus expression system. *J. Biol. Chem.*, 280:3208–3216, **2005**.

Shental-Bechor, D.; Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A.*, 105:8256-8261, **2008**.

Shields, R. L.; Lai, J.; Keck, R.; O'Connell, L. Y.; Hong, K.; Meng, Y. G.; Weikert, S. H.; Presta, L. G. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J. Biol. Chem.*, 277:26733–26740, **2002**.

Shinkawa, T.; Nakamura, K.; Yamane, N.; Shoji-Hosaka, E.; Kanda, Y.; Sakurada, M.; Uchida, K.; Anazawa, H.; Satoh, M.; Yamasaki, M.; Hanai, N.; and Shitara, K. The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *J Biol Chem.*, 278:3466–3473, **2003**.

Shriver, Z.; Raguram, S.; and Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nat. Rev. Drug Disc.*, 3:863–873, **2004**.

Skropeta, D. The effect of individual N-glycans on enzyme activity. *Bioorg. Med. Chem.*, 17:2645-2653, **2009**.

Sperandio, M.; Thatte, A.; Foy, D.; Ellies, L. G.; Marth, J. D.; and Ley, K. Severe impairment of leukocyte rolling in venules of core 2 glucosaminyltransferase-deficient mice. *Blood*, 97:3812–3819, **2001**.

Spiro, R. G.: Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12:43R-56R, **2002**.

Stewart, J. J. P. Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations. *Int. J. Quant. Chem.*, 58:133-146, **1996**.

Stewart, J. J. P., Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Modeling*, 13:1173-1213, **2007**.

Stewart, J. J. P. MOPAC2009, Stewart Computational Chemistry, Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net), **2008**.

Taketa, K.; Endo, Y.; Sekiya, C.; Tanikawa, K.; Kojii, T.; Taga, H.; Satomura, S.; Matsuura, S.; Kawai, T.; and Hirai, H. A collaborative study for the evaluation of lectin-reactive α-fetoproteins in early detection of hepatocellular carcinoma. *Cancer Res.*, 53:5419–5423, **1993**.

Tams, J. W.; Vind, J.; and Welinder, K. G. Adapting protein solubility by glycosylation. N-glycosylation mutants of *Coprinus cinereus* peroxidase in salt and organic solutions. *Biochim Biophys Acta*. 1432:214-221, **1999**.

Taniguchi, N.; Honke, K.; and Fukuda, M. Eds., Springer, TokyoBerlinHeidelbergNew YorkLondon, **2002** – Book Review: Biochemistry (Moscow), 68:707-708, **2003**. Translated from Biokhimiya, 68:861-862 2003, Original Russian Text Copyright © 2003 by Wiederschain.

Taniguchi, N.; Miyoshi, E.; Jianguo, G.; Honke, K.; and Matsumoto, A. Decoding sugar functions by identifying target glycoproteins. *Curr Opin Struct Biol.*, 16:561–566, **2006**.

Tarbouriech, N.; Charnock, S. J.; and Davies, G. J. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. *J. Mol. Biol.*, 314:655–661, **2001**.

Turner, G. A. N-Glycosylation of serum-proteins in disease and its investigation using lectins. *Clin. Chim. Acta*, 208:149–171, **1992**.

Vailes, L. D.; Li, Y.; Bao, Y.; DeGroot, H.; Aalberse, R. C.; and Chapman, M. D. *J. Allergy. Clin. Immunol.*, 93,22–33, **1994**.

van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; and Berendsen, H. J. C. GROMACS: fast, flexible and free. *J. Comput. Chem.*, 26:1701-1718, **2005**.

van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Huenenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.: Biomolecular Simulation: The GROMOS96 Manual and User Guide. Vdf Hochschulverlag AG, Zuriq, Suíça, **1996**.

Vanhooren, V.; Dewaele, S.; Kuro-o, M.; Taniguchi, N.; Dolle, L.; van Grunsven, L. A.; Makrantonaki, E.; Zouboulis, C. C.; Chen, C. C.; and Libert, C. Alteration in N-glycomics during mouse aging: a role for FUT8. *Aging Cell*, 10:1056–1066, **2011**.

Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3:97-130, **1993**.

Varki, A. In *Essentials of Glycobiology* (eds Varki, A. et al.) 537–549 (Cold Spring Harbor Laboratory Press, New York), **1999**.

Varki, A., Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature*, 446:1023–1029, **2007**.

Verli, H.: Níveis de informação biológica. In Verli, H (Ed.) *Bioinformática: da biologia molecular à flexibilidade molecular*. Porto Alegre: e-book (<http://www.ufrgs.br/bioinfo/ebook/>), 2014.

Wang, X.; Inoue, S.; Gu, J.; Miyoshi, E.; Noda, K.; Li, W.; Mizuno-Horikawa, Y.; Nakano, M.; Asahi, M.; Takahashi, M.; Uozumi, N.; Ihara, S.; Lee, S. H.; Ikeda, Y.; Yamaguchi, Y.; Aze, Y.; Tomiyama, Y.; Fujii, J.; Suzuki, K.; Kondo, A.; Shapiro, S. D.; Lopez-Otin, C.; Kuwaki, T.; Okabe M.; Honke, K.; and Taniguchi, N. Dysregulation of TGF- β 1 receptor activation leads to abnormal lung development and emphysemalike phenotype in core fucose-deficient mice. *Proc Natl Acad Sci U S A.*, 102:15791–15796, **2005**.

Wang, X.; Gu, J.; Ihara, H.; Miyoshi, E.; Honke, K.; Taniguchi, N. Core fucosylation regulates epidermal growth factor receptor-mediated intracellular signaling. *J Biol Chem.*, 281:2572–2577, **2006**.

Wang, X.; Fukuda, T.; Li, W.; Gao, C.; Kondo, A.; Matsumoto, A.; Miyoshi, E.; Taniguchi, N.; and Gu, J. Requirement of Fut8 for the expression of vascular endothelial growth factor receptor-2: a new mechanism for the emphysema-like changes observed in Fut8-deficient mice. *J. Biochem.*, 145:643–651, **2009**.

Williams, D.; Longmore, G.; Matta, K. L.; and Schachter, H. Mucin synthesis. II. Substrate specificity and product identification studies on canine submaxillary gland UDP-GlcNAc:Gal beta 1-3GalNAc(GlcNAc leads to GalNAc) beta 6-N-acetylglucosaminyltransferase. *J. Biol. Chem.*, 255:11253–11261, **1980**.

Wlodawer, A.; Minor, W.; Dauter, Z.; and Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.*, 275:1-21, **2008**.

Woods, R. J. Computational carbohydrate chemistry: what theoretical methods can tell us. *Glycoconj. J.*, 15:209-216, **1998**.

Wormald, M. R.; Petrescu, A. J.; Pao, Y. -L.; Glithero, A.; Elliott, T.; and Dwek, R. A. Conformational Studies of Oligosaccharides and Glycopeptides: Complementarity of NMR, X-ray Crystallography, and Molecular Modelling. *Chem. Rev.*, 102:371-386, **2002**.

Wyss, D. F.; and Wagner, G. The structural roles of sugars in glycoproteins. *Curr. Opin, Biotechnol.*, 7:409-416, **1996**.

Yeh, J. C.; Ong, E.; and Fukuda, M. Molecular Cloning and Expression of a Novel β -1,6-N-Acetylglucosaminyltransferase That Forms Core 2, Core 4, and I Branches. *J. Biol. Chem.*, 274:3215–3221, **1999**.

Zhao, Y.; Itoh, S.; Wang, X.; Isaji, T.; Miyoshi, E.; Kariya, Y.; Miyazaki, K.; Kawasaki, N.; Taniguchi, N.; and Gu, J. Deletion of core fucosylation on α 3b1 integrin down-regulates its functions. *J Biol Chem.*, 281:38343–38350, **2006**.

9. APÊNDICES

Apêndice A

Capítulo de livro: Estrutura de receptores-alvo de fármacos
através de modelagem comparativa

Sachett LG, Ligabue-Braun R, Verli H

In Trossini G, Castilho MS (Eds.), *Práticas de Modelagem Molecular*

O capítulo a seguir, apresentado em sua formatação preliminar, é parte do livro-texto “Práticas de Modelagem Molecular” que encontra-se em fase final de preparação. Algumas seções do capítulo (como práticas e exercícios avançados) foram omitidas da presente tese por solicitação dos editores.

ESTRUTURA DE RECEPTORES-ALVO DE FÁRMACOS ATRAVÉS DE MODELAGEM COMPARATIVA

Liana G. Sachett, Rodrigo L. Braun, Hugo Verli.

Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia

Resumo

Uma das principais dificuldades associadas ao planejamento de novos fármacos está na ausência de informações sobre a estrutura 3D de seus receptores-alvo. Uma das técnicas computacionais disponíveis para contornar esta limitação é denominada modelagem comparativa. Esta técnica consiste na construção de um modelo 3D para uma dada proteína usando, como referência (ou molde), a estrutura de uma proteína semelhante. O modelo assim obtido poderá ser empregado em outros estudos, como ancoramento (docking). Os passos necessários para a realização da modelagem comparativa serão descrito em detalhes neste capítulo, incluindo-se exemplos, limitações e dificuldades mais comuns, bem como interpretação de resultados.

INTRODUÇÃO

Diversas estratégias de desenvolvimento de novos fármacos estão baseadas no conhecimento da estrutura 3D do receptor-alvo, tais como o planejamento baseado na estrutura do receptor (structure-based drug design), estudos de ancoramento (docking) e simulações por dinâmica molecular. Adicionalmente, métodos como QSAR e SAR também podem empregar informações sobre a estrutura 3D da proteína-alvo. Infelizmente, em torno de 3% das proteínas humanas tiveram suas estruturas 3D determinadas experimentalmente, o que demanda o emprego de técnicas adicionais.

Uma das principais técnicas computacionais empregada na obtenção de estruturas 3D de proteínas, tais como receptores-alvo de fármacos, é denominada modelagem comparativa de proteínas ou modelagem por homologia. Esta técnica é baseada no pressuposto de que proteínas com sequências de aminoácidos semelhantes possuem estruturas 3D semelhantes. Esta semelhança nas sequências representa, usualmente, uma origem evolutiva comum, o que é denominado homologia. Assim, a proteína de interesse do Químico Medicinal, para a qual não há estrutura 3D disponível (denominada de sequência-alvo), deve ser usualmente homóloga a uma proteína que possua estrutura 3D já estabelecida (denominada de molde) para que a técnica de modelagem por homologia possa ser empregada.

Contudo, o conceito de homologia implica em ser ou não ser homólogo, nos impedindo de quantificar o “grau de homologia”. Isto é importante, porque as sequências de aminoácidos de duas proteínas podem ser muito ou pouco parecidas. E quanto mais “parecidas” forem estas sequências (da sequência-alvo e do molde), maior será a confiabilidade do modelo obtido. Uma das

formas de comparar duas sequências de aminoácidos é através do grau de identidade entre as mesmas, de 0 a 100%. Em outras palavras, a identidade mede o número de resíduos de aminoácidos que são idênticos e ocupam as mesmas posições nas sequências das proteínas em estudo. A identidade mínima entre a sequência-alvo e o molde deverá ser de 30% para obtenção de modelos confiáveis por modelagem comparativa. Quanto maior o grau de identidade, mais confiável será o modelo.

Usualmente, a modelagem comparativa compreende os seguintes passos: busca por moldes, seleção de um ou mais moldes, construção do modelo e avaliação do modelo (Figura 1). Adicionalmente, embora diversas ferramentas computacionais estejam disponíveis para a realização de estudos de modelagem comparativa, empregaremos aqui o servidor denominado SWISS-MODEL (<http://swissmodel.expasy.org/>). Considere que algumas pequenas diferenças nos procedimentos apresentados podem ser encontradas em outros programas, embora o roteiro geral seja bastante semelhante.

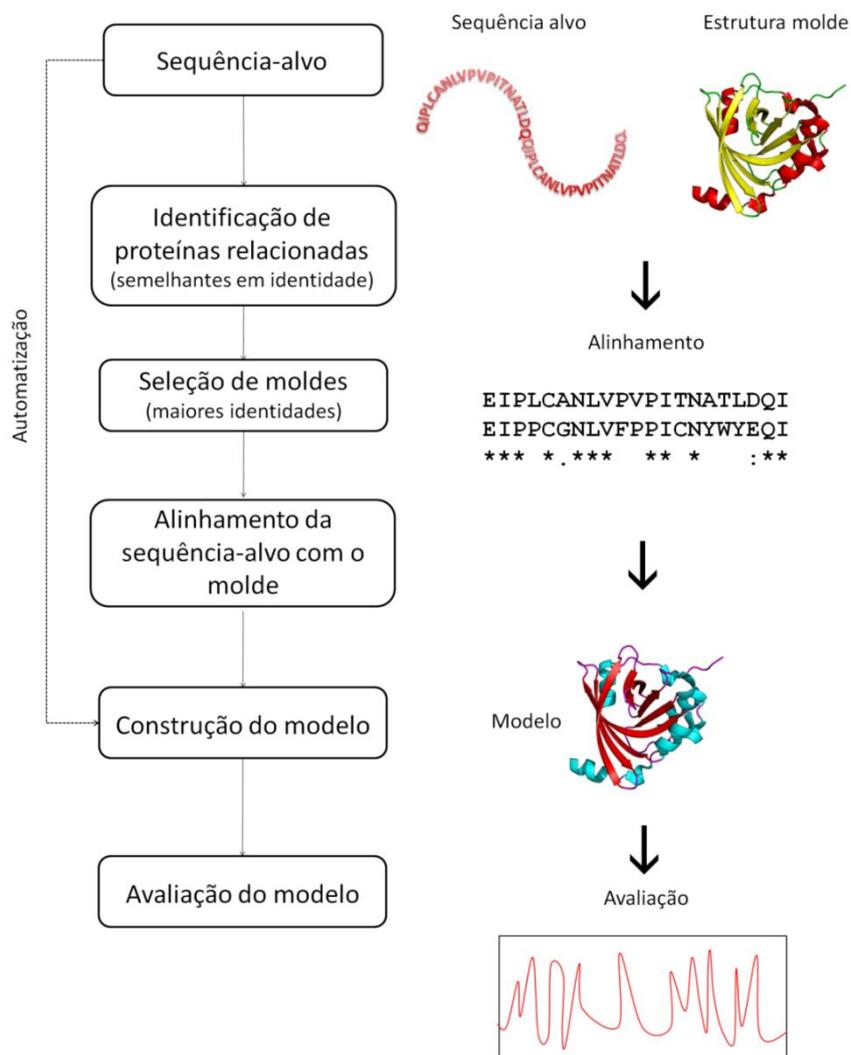


Figura 1. Passos da modelagem comparativa (baseado em Sánchez & Šali, 2000).

A identificação de estruturas relacionadas (Figura 1) é realizada comparando (ou alinhando) a sequência-alvo à sequências de aminoácidos de proteínas que já tenham estruturas 3D determinadas. Esta etapa pode ser realizada, por exemplo, utilizando-se a ferramenta BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), selecionando o menu “protein blast”. Gera-se, assim, uma lista de diferentes proteínas, relacionadas à sequência-alvo, candidatos a moldes para a modelagem comparativa. Adicionalmente, são apresentadas as identidades entre a sequência-alvo e cada possível molde, além de seus respectivos alinhamentos.

O alinhamento realizado entre a sequência-alvo e seu molde é uma das partes mais importantes da modelagem comparativa de proteínas. A definição de quais resíduos de aminoácidos correspondem entre as duas macromoléculas torna-se determinante da forma 3D a ser adotada pela sequência-alvo. Por outro lado, a seleção de um molde, a partir dos candidatos obtidos pelo alinhamento, tem como regra mais simples a seleção da proteína com a maior identidade em relação à sequência-alvo. Adicionalmente, deve-se considerar a qualidade da estrutura 3D experimental, por exemplo pela resolução cristalográfica, medida em angstroms (Å). Quanto menor o valor da resolução, melhor (idealmente em torno de 2,5 Å ou menor).

A etapa final da modelagem comparativa consiste na validação do modelo, onde diversas características, principalmente estereoquímicas e funcionais, são avaliadas. Neste processo, a proteína modelada deve tanto apresentar uma estrutura 3D satisfatória, seguindo alguns critérios previamente estabelecidos pela comunidade científica, quanto ser capaz de explicar aspectos funcionais já conhecidos, como aqueles relacionados à catálise (como a composição e geometria da tríade catalítica), à

ligação de cofatores estruturais (como metais) ou de moduladores (neurotransmissores ou fármacos, por exemplo). A qualidade do modelo dependerá, principalmente, da qualidade do molde escolhido e do alinhamento. A quantidade de alças na proteína, alinhamentos contendo erros ou gaps muito longos são alguns fatores que também podem influenciar na qualidade do modelo obtido.

Diversas estratégias estão disponíveis para realizar a validação estereoquímica dos modelos obtidos. Uma das principais envolve o Mapa de Ramachandran. Resumidamente, este mapa apresenta geometrias (ou zonas) proibidas e permitidas para o esqueleto peptídico da proteína. Para a obtenção deste dado pode ser utilizado, por exemplo, o programa PROCHEK através do servidor PDBSUM (www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html).

Outras estratégias comuns para validação estereoquímica podem ser encontradas na plataforma SWISS-MODEL e incluem o QMEAN, que pontua a confiabilidade do modelo a partir de comparações à estruturas cristalográficas de alta resolução, e a Análise de Erros Locais, que consiste em estimar a qualidade estereoquímica resíduo a resíduo e entre resíduos vizinhos. Este dado é apresentado na forma de um gráfico contendo pontuações para cada resíduo que refletem a probabilidade de erro em cada um deles. Para facilitar a interpretação dos resultados, uma representação do modelo proteico colorido por esta pontuação de erros é apresentada simultaneamente.

A validação funcional dos modelos, ao contrário da estereoquímica, é baseada mais na análise da literatura científica do que em resultados de programas ou servidores na internet. Em geral, busca-se o maior volume possível de informações bioquímicas, estruturais, funcionais e terapêuticas disponíveis, e avalia-se a capacidade do

modelo em responder ou explicar estas questões.

A modelagem comparativa pode ser realizada tanto de uma forma automatizada, frequentemente disponível em servidores na internet, quanto de uma forma manual, usualmente baseada em programas que funcionam localmente, no computador do usuário. Nos casos mais difíceis, o emprego de estratégias manuais pode ser necessária, maximizando a qualidade dos modelos. Para estes casos mais delicados, ou quando o usuário deseja um controle total de todas as etapas realizadas, o programa MODELLER é um dos mais utilizados (www.salilab.org/modeller). Na página do programa encontram-se instruções para instalação e tutoriais para o seu uso.

Discussão e Conclusão

Ao realizar os passos essenciais na construção de um modelo tridimensional de proteínas, a atividade proposta no roteiro de aula prática evidencia tanto a praticidade quando a aplicabilidade do método. Como observado ao longo do capítulo, a maioria das etapas não apresenta grandes dificuldades em sua execução, podendo ser seguida à maneira de um protocolo. Em casos de modelagem manual, a escolha da estrutura molde, o alinhamento das sequências alvo e molde e a seleção de parâmetros para construção do modelo são realizados pelo usuário. No procedimento automatizado, uma vez de posse da sequência polipeptídica, o próprio servidor de modelagem seleciona o melhor molde, o alinha à sequência-alvo e executa as rotinas de construção tridimensional. Ao seguir uma abordagem automatizada para modelagem, o ponto crítico passa a ser a avaliação do modelo obtido. A qualidade e a validade do modelo poderão ser definidas apenas por verificação manual dos resultados de sua avaliação, caracterizando o modelo obtido como realístico e confiável. De fato,

esta verificação é essencial para que estudos avaliando a interação entre ligante e receptor possam ser realizados de forma realística. De maneira geral, modelos ruins podem levar a conclusões ruins.

Entre as avaliações essenciais está o gráfico de Ramachandran. De interpretação objetiva, este gráfico permite tanto avaliar a qualidade global do modelo, quanto identificar resíduos específicos que possam apresentar falhas estereoquímicas. Adicionalmente, do ponto de vista funcional, é de grande importância avaliar a integridade do sítio catalítico na estrutura modelada, no caso de enzimas, como a DOPAd, enzima exemplificada no roteiro de aula prática. Isso é geralmente feito por comparação com dados da literatura. Proteínas não-enzimáticas (como carreadores de fármacos, canais iônicos ou receptores acoplados à proteína G) também apresentam resíduos importantes para sua atividade, e estes devem ser inspecionados no modelo obtido.

Após avaliação e validação, pode-se verificar a alta qualidade da estrutura da DOPA descarboxilase humana obtida nesta atividade. Tal estrutura pode ser empregada em estudos de planejamento racional de novos fármacos, por exemplo, através de ancoramento (docking) molecular. Ainda, a estrutura obtida por modelagem comparativa pode ser visualizada em softwares como PyMol (Capítulo XX), VMD, UCSF Chimera e Swiss PDB Viewer.

Ao realizar a atividade de modelagem comparativa baseada na DOPAd, são desenvolvidos os passos essenciais na construção de um modelo tridimensional de uma proteína. Este é, contudo, um exemplo geral, existindo casos especiais na construção de estruturas tridimensionais. Alguns deles são discutidos nos exercícios adicionais (proteínas transmembrana e acopladas a membranas). Em outros casos, entretanto, não é possível aplicar a modelagem comparativa. Tais casos

geralmente envolvem proteínas para as quais não existem moldes disponíveis. Uma discussão sobre estes casos está além do objetivo deste capítulo, sendo o leitor encaminhado ao artigo de Jones (2000) como um guia inicial para estes casos. Apesar de poder ser empregada isoladamente, em busca da estrutura de uma proteína específica, a modelagem comparativa constitui uma etapa intermediária em estudos de desenvolvimento de fármacos, auxiliando na compreensão das relações estrutura-atividade de compostos bioativos, utilizando ferramentas gratuitas e dados disponíveis publicamente.

REFERÊNCIAS

- CHOTHIA, C.; LESK, A.M. The relation between the divergence of sequence and structure in proteins. *EMBO J.*, 1986, 5, 823-826.
- DE BREVERN, A.G. 3D structural models of transmembrane proteins. *Methods Mol. Biol.*, 2010, 654, 387-401.
- FISER, A.; DO, R.K.G.; ŠALI, A. Modeling of loops in protein structures. *Protein Sci.*, 2000, 9, 1753-1773.
- FORSTER, M.J. Molecular modelling in structural biology. *Micron*, 2002, 33, 365-384.
- JONES, D.T. A practical guide to protein structure prediction. *Methods Mol. Biol.*, 2000, 143, 131-154.
- LESK, A.M. Estrutura de proteínas e descoberta de fármacos. In: _____. *Introdução à Bioinformática*. Porto Alegre: Artmed, 2008.
- MARTÍ-RENOM, M.A.; et al. Comparative protein structure modeling of genes and genomes. *Annu. Rev.. Biophys. Biomol. Struct.*, 2000, 29, 291-325.
- REECK, G.R. et al. "Homology" in protein and nucleic acids: a terminology muddle and a way out of it. *Cell*, 1987, 50, 667.
- ŠALI, A; BLUNDELL, T.L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, 1993, 234, 779-815.
- SÁNCHEZ, R.; ŠALI, A. Comparative protein structure modeling. Introduction and practical examples with Modeller. *Methods Mol. Biol.*, 2000, 143, 97-129.
- SANTOS FILHO, O.A.; ALENCASTRO, R.B. Modelagem de proteínas por homologia. *Quim. Nova*, 2003, 26, 253-259.
- TRAMONTANO, A. The role of molecular modelling in biomedical research. *FEBS Lett.*, 2006, 580, 2928-2934.

Apêndice B

**Artigo: Conformational Characterization of Ipomotaosides and their
recognition by COX-1 and 2**

Pablo R Arantes, Liana G Sachett, Cedric S Graebin, Hugo Verli

Molecules 2014, 19, 5421-5433.

Article

Conformational Characterization of Ipomotaosides and Their Recognition by COX-1 and 2

Pablo R. Arantes ¹, Liana G. Sachett ¹, Cedric S. Graebin ² and Hugo Verli ^{1,*}

¹ Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500, CP 15005, Porto Alegre, RS 91500-970, Brazil; E-Mails: pabloarantes@cbiot.ufrgs.br (P.R.A.); lianasachett@cbiot.ufrgs.br (L.G.S.)

² Departamento de Química, Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Cidade Universitária, Seropedica, RJ 23897-000, Brazil; E-Mail: cedric@ufrj.br

* Author to whom correspondence should be addressed; E-Mail: hverli@cbiot.ufrgs.br; Tel./Fax: +55-51-3308-7770.

Received: 5 March 2014; in revised form: 18 April 2014 / Accepted: 21 April 2014 /

Published: 24 April 2014

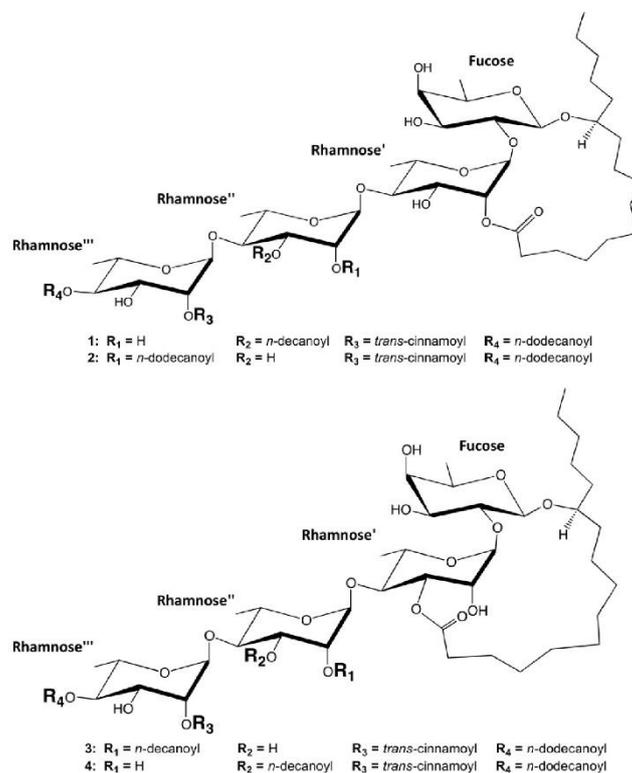
Abstract: The aerial parts of *Ipomoea batatas* are described herein to produce four new resin glycosides, designated as ipomotaosides A, B, C, and D. Ipomotaoside A was found to present inhibitory activity on both cyclooxygenases. However, the conformational elucidation of these molecules may be difficult due to their high flexibility. In this context, the current work presents a conformational characterization of ipomotaosides A–D in aqueous and nonaqueous solvents. The employed protocol includes metadynamics evaluation and unrestrained molecular dynamics simulations (MD). The obtained data provided structural models for the ipomotaosides in good agreement with previous ROESY distances measured in pyridine. Accordingly, the most abundant conformation of ipomotaoside A in solution was employed in flexible docking studies, providing a structural basis for the compound's inhibition of COX enzymes. The so-obtained complex supports resin glycosides' role as original scaffolds for future studies, aiming at structural optimization and development of potential new anti-inflammatory agents.

Keywords: ipomotaosides; resin glycosides; disaccharides; molecular dynamics; docking; COX; inflammatory process

1. Introduction

Ipomotaosides are resin glycosides derived from the aerial parts of *Ipomea batatas*. These resin glycosides are separated in four different structures, named ipomotaosides A (1), B (2), C (3), and D (4) [1] (Figure 1).

Figure 1. Representations of the ipomotaosides characterized in the current work.



These compounds are composed of a macrocyclic structure, various hydrophobic acyl chains and the same hydrophilic oligosaccharide core. Differently from the roots of *I. batatas*, which are used in folkloric medicine for anemia, diabetes, hemorrhage and hypertension [2], resin glycosides from the aerial parts as well as roots are inactive cytotoxic agents and known to exhibit several biological effects such as inhibition of multidrug resistance efflux pumps (EPIs) [3–5]. Additionally, recent results demonstrated that ipomotaoside A from *I. batatas* is capable of inhibiting cyclooxygenases (COX) 1 and 2 [1]. These enzymes are committed in prostanoid biosynthesis, converting arachidonic acid (AA) and O₂ to prostaglandin endoperoxide PGH₂ in two different active sites, cyclooxygenase and peroxidase [6]. The first step consists in oxygenating AA to prostaglandin G₂ (PGG₂) in the cyclooxygenase site. In the second step, this intermediate moves to the peroxidase active site to be

reduced to PGH_2 , which is converted into other prostaglandins and thromboxanes responsible for mediating the inflammatory process [6], so the inhibition of both COXs by ipomotaoside A indicates a potential role of these molecules in modulating inflammation.

Future efforts to develop these compounds through rational design, however, are impaired by the lack of structural information on their complexation to COX. Although there are crystallized glycoconjugates described in the literature [7] due to the high flexibility of glycosidic linkages in carbohydrates [8,9] and acyl chains and the consequent high amount of conformers coexisting in solution, [8,10,11] it is unusual to obtain 3D models for glycoconjugates from crystallographic methods. On the other hand, NMR methods are the main choice for dealing with such flexible compounds, for example, if a reasonable amount of ROESY/NOESY contacts are obtained [8]. Still, the chemical environment in which NMR experiments are performed may not correspond to that of physiological solutions, with potential conformational influences. In the case of ipomotaosides, the spectroscopic elucidation was performed on pyridine, a heterocyclic and aromatic solvent that does not offer a good resemblance to biological media.

Considering the adversities found in obtaining atomistic models for complex carbohydrates and glycoconjugates in environments mimicking biological solutions, the current work aimed to characterize ipomotaosides 1–4 in aqueous and pyridine solutions. Therefore, the conformational characterization of all molecules was included in the manuscript in order to offer to the researcher of the field additional structural information on this class of compounds. Also, the conformational ensemble in pyridine was used as reference for spectroscopic validation, and the so-obtained most abundant conformation of ipomotaoside A in aqueous solution was submitted to docking studies to provide insights into its inhibitory activity against COXs.

The strategy employed for conformational characterization of ipomotaosides was previously described [11] and validated against NMR data for compounds such as saponins [12], exopolysaccharides [13], galactans and fucans [14], and a series of glycoproteins and glycopeptides [11,15], based on building glycan chains from most abundant conformational states in solution, as determined by MD simulations. The built molecules were submitted to additional simulations in order to account to potential inter-residue interactions and, consequently, conformational effects.

2. Results and Discussion

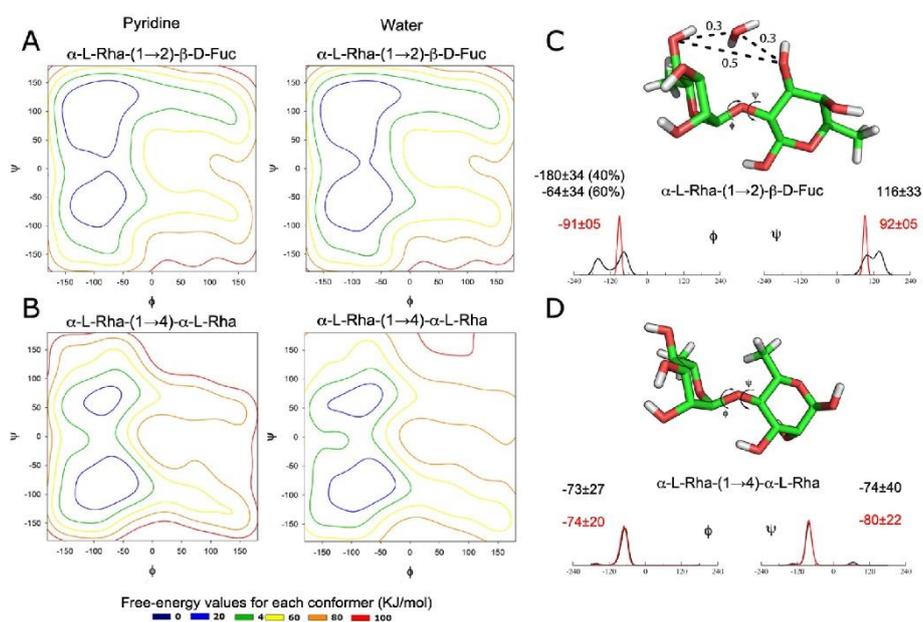
2.1. Dynamics of Isolated Disaccharides

On the basis of the ipomotaosides' structures, two glycosidic linkages had their conformational behavior were evaluated by metadynamics, in both pyridine and water: $\alpha\text{-L-Rha-(1}\rightarrow\text{4)-}\alpha\text{-L-Rha}$ and $\alpha\text{-L-Rha-(1}\rightarrow\text{2)-}\beta\text{-D-Fuc}$. The obtained patterns point to similar conformations in both solvents, with a displacement of the global minimum from the northwest to the southwest quadrants upon modification of rhamnose residue linkage from C-2 (fucose) to C-4 (rhamnose) (Figure 2A,B).

The obtained conformational behavior of ipomotaosides disaccharide units was further compared to previous crystallographic and molecular mechanics (MM3) data of tricolorin A, the only member of this class of oligosaccharides with a published crystal structure [7] for the $\alpha\text{-(1}\rightarrow\text{2)}$ linkage. The two main minima are located in similar regions as clearly demonstrated for both works. However, the main

minimum is inverted, located in the southwest quadrant in MM3 (region in which the crystallographic geometry was observed), and in the northwest quadrant in GROMOS. Some aspects may be related to this difference, such as the nature of the force field (united atom or all atom, validated against condensed or gas phase) and of the employed method (metadynamics or adiabatic maps). Also, it should be noted that the location of the minimum in pyridine is on the opposite quadrant than the crystallographic geometry.

Figure 2. (A,B) Conformational behavior of α -L-Rha-(1 \rightarrow 2)- β -D-Fuc and α -L-Rha-(1 \rightarrow 4)- α -L-Rha linkages, as obtained from metadynamics in both pyridine and water; (C,D) Distribution of Φ and Ψ dihedral angles during simulations associated with the glycosidic linkages of isolated disaccharides. The aqueous solution is represented as black and the nonaqueous solution (pyridine) is indicated as red.



Despite the identification of two minima on the metadynamics contour plots, the MD simulations of both minima populated a single region, corresponding to the global minima (Figure 2C,D), which could indicate a similar conformational state between the two solvents. Still, glycosidic linkages presented a more rigid pattern on pyridine than on water, in agreement with the usual choice of the first for spectroscopic studies. Even being more rigid, the linkages dynamic under pyridine lies within the conformational behavior observed in water (Figure 2C,D). This data is in agreement with previous results indicating that pyridine has a discrete effect on molecule conformation [12].

The conformation observed in water for Rha-(1→2)-β-D-Fuc, with the appearance of a second minimum at the Φ angle, is a result of the hydrogen bonds between water and hydroxyl groups of of rhamnose C-4 and fucose C-3 (Figure 2C). This behavior is demonstrated at radial distribution function (Figure S1), where the respective oxygen atoms indicate the presence of a water molecule during the simulation. The distance between these two oxygen atoms is lower when the water is present (Figures 2C and S2), showing a solvent influence on Φ angle and so resulting on its second minimum in water.

The most abundant conformational states of each glycosidic linkage were then used as starting geometries for the construction of complete models of structures 1–4, previously described as a successful approach to obtain 3D models of glycan chains in solution, [11–13,15] and submitted to MD simulations for further geometry refinement in the whole molecule scaffold as well as for conformational sampling.

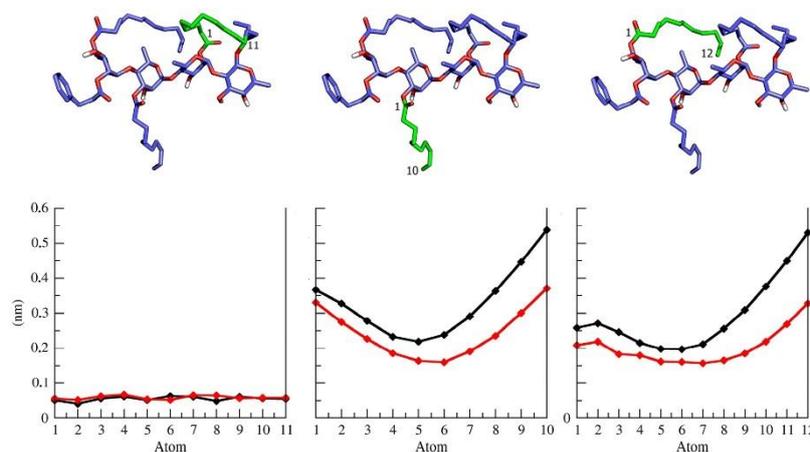
2.2. Ipomotaosides Dynamics

The MD simulation of complete ipomotaosides allowed a comparison between the conformational profile adopted by the isolated glycosidic linkages and their behavior when composing the whole molecules (Table 1 and Figure S3). For the α -L-Rha-(1→2)-β-D-Fuc linkage, upon inclusion within ipomotaosides in pyridine, a new conformational state, at Φ and Ψ angles, is observed in compounds 1, 3, and 4 (Table 1) in comparison to the isolated linkage (Figure 2). Such effect appears to be related to the rigidity promoted by the macrocycle, more pronounced than the influence of acyl chains (Figure 3 and Figures S4–S6).

Table 1. Dihedral angles from Ipomotaosides glycosidase linkages, as obtained from MD.

Compound	Condition	Glycosidic Linkage					
		α -L-Rha'-(1→2)- β-D-Fuc		α -L-Rha''-(1→4)- α-L-Rha'		α -L-Rha'''-(1→4)- α-L-Rha''	
		Φ	Ψ	Φ	Ψ	Φ	Ψ
Ipomotaoside 1	MD in pyridine	-171 ± 29 (42%) -73 ± 28 (58%)	79 ± 40 (37%) -51 ± 29 (63%)	-75 ± 22 -	-85 ± 14 -	-71 ± 11 -	-81 ± 10 -
	MD in water	-163 ± 15	86 ± 19	-74 ± 22	-87 ± 21	-71 ± 15	-82 ± 12
Ipomotaoside 2	MD in pyridine	-165 ± 14	82 ± 15	-76 ± 14	-83 ± 11	-76 ± 14	-83 ± 11
	MD in water	-71 ± 26 (14%) -170 ± 31 (86%)	-51 ± 13 (14%) 81 ± 56 (86%)	-76 ± 42 -	-87 ± 19 -	-78 ± 40 -	-88 ± 18 -
Ipomotaoside 3	MD in pyridine	-183 ± 25 (47%) -91 ± 29 (53%)	76 ± 27 (47%) -35 ± 37 (53%)	-74 ± 11 -	-82 ± 10 -	-75 ± 14 -	-82 ± 11 -
	MD in water	-173 ± 10	87 ± 11	-72 ± 26	-82 ± 13	-77 ± 37	-86 ± 17
Ipomotaoside 4	MD in pyridine	-171 ± 25 (43%) -85 ± 25 (57%)	83 ± 26 (43%) -31 ± 40 (57%)	-71 ± 39 -	-85 ± 16 -	-74 ± 11 -	-84 ± 10 -
	MD in water	-173 ± 09 -	87 ± 09 -	-97 ± 33 (34%) -185 ± 18 (66%)	-83 ± 20 (34%) -150 ± 18 (66%)	-80 ± 26 -	-84 ± 15 -

Figure 3. Root mean square fluctuation (RMSF) for ipomotaoside A macrocycle and acyl chains. A tridimensional representation of each structure is highlighted in green, at the top, and the corresponding RMSF is at the bottom. Aqueous solution is represented in black and pyridine is indicated in red.



In fact, a similar situation is described in the literature for tricolorin A [7], a glycoconjugate in which the macrocycle imposes a conformational limit on the linked carbohydrate residues. On the other hand, the α -L-Rha-(1 \rightarrow 4)- α -L-Rha linkage in pyridine and water, outside the macrocycle, shows no conformational dependence on the entire molecule scaffold (Table 1 and Figure S3), pointing again to a minor conformational influence of the hydrophobic acyl chains, as demonstrated on tricolorin A [7].

Considering that structures 1–4 have been previously characterized by NMR spectroscopy, [1] the interproton contacts were used to validate the conformational ensemble obtained from unrestrained MD simulations (Table 2). As a general feature, most of the experimentally observed contacts between ipomotaosides protons were properly reproduced on the performed simulations, pointing to a precise conformational characterization of these compounds in pyridine. These results are in agreement with previous Calonyctin A restrained MD studies [16], another glycoconjugate of the resin glycoside type, where the average of NOESY violation is less than 10%. Such results also supported the validity of the conformational sampling of these compounds in water, since no spectroscopic data is available for this solvent.

While ipomotaoside A represents a new scaffold to modulate COX [1], future medicinal chemistry efforts to guide the optimization of its chemical structure may be reinforced by the elucidation of the inhibitor-enzyme complex 3D structure. Such a complex could also explain the molecular basis for the activity of these structurally unusual compounds when compared to drugs in clinical use modulating this enzyme. In this context, the most abundant conformational states of ipomotaoside A in water were employed in docking studies on both COX enzymes.

Table 2. Comparison between ROESY * contacts of Ipomotaosides and the interproton distances from MD Simulations.

Ipomotaoside	Proton 1	Proton 2	Average of Interproton distance from MD (Å)	Ipomotaoside	Proton 1	Proton 2	Average of Interproton distance from MD (Å)
1	Fuc-(H1)	Fuc-(H5)	2.4 ± 0.2	3	Fuc-(H1)	Fuc-(H5)	2.4 ± 0.2
	Rha ¹ -(H1)	Fuc-(H1)	2.5 ± 0.7		Rha ¹ -(H1)	Fuc-(H1)	2.5 ± 0.7
	Rha ¹ -(H1)	Fuc-(H2)	3.2 ± 0.5		Rha ¹ -(H1)	Fuc-(H2)	3.2 ± 0.4
	Rha ¹ -(H1)	Fuc-(H3)	3.4 ± 0.8		Rha ¹ -(H1)	Fuc-(H3)	3.6 ± 0.8
	Rha ¹ -(H1)	Fuc-(H4)	5.5 ± 0.5		Rha ¹ -(H5)	Fuc-(H1)	4.9 ± 0.3
	Rha ¹ -(H1)	Rha ¹ -(H3)	3.4 ± 0.3		Rha ¹ -(H1)	Rha ¹ -(H2)	4.7 ± 0.2
	Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.3		Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.2
	Rha ¹ -(H1)	Rha ¹ -(H4)	2.6 ± 0.3		Rha ¹ -(H1)	Rha ¹ -(H5)	4.4 ± 0.1
2	Fuc-(H1)	Fuc-(H5)	2.4 ± 0.2	Rha ¹ -(H1)	Rha ¹ -(H3)	3.3 ± 0.2	
	Rha ¹ -(H1)	Fuc-(H1)	3.3 ± 0.3	Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.3	
	Rha ¹ -(H1)	Fuc-(H2)	2.6 ± 0.3	Rha ¹ -(H1)	Rha ¹ -(H5)	4.4 ± 0.1	
	Rha ¹ -(H1)	Fuc-(H3)	4.4 ± 0.1	4	Fuc-(H1)	Fuc-(H5)	2.4 ± 0.2
	Rha ¹ -(H2)	Fuc-(H2)	2.2 ± 0.3		Rha ¹ -(H1)	Fuc-(H1)	2.5 ± 0.7
	Rha ¹ -(H2)	Fuc-(H3)	4.7 ± 0.3		Rha ¹ -(H1)	Fuc-(H2)	3.1 ± 0.5
	Rha ¹ -(H1)	Rha ¹ -(H3)	3.4 ± 0.2		Rha ¹ -(H1)	Fuc-(H3)	3.5 ± 0.8
	Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.3		Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.3
	Rha ¹ -(H1)	Rha ¹ -(H5)	4.4 ± 0.1		Rha ¹ -(H1)	Rha ¹ -(H5)	4.3 ± 0.2
	Rha ¹ -(H2)	Rha ¹ -(H3)	4.1 ± 0.3		Rha ¹ -(H1)	Rha ¹ -(H4)	2.5 ± 0.3
	Rha ¹ -(H2)	Rha ¹ -(H4)	4.3 ± 0.2		Rha ¹ -(H1)	Rha ¹ -(H5)	4.4 ± 0.1
	Rha ¹ -(H1)	Rha ¹ -(H3)	3.4 ± 0.2	-	-	-	
	Rha ¹ -(H2)	Rha ¹ -(H3)	4.1 ± 0.3	-	-	-	

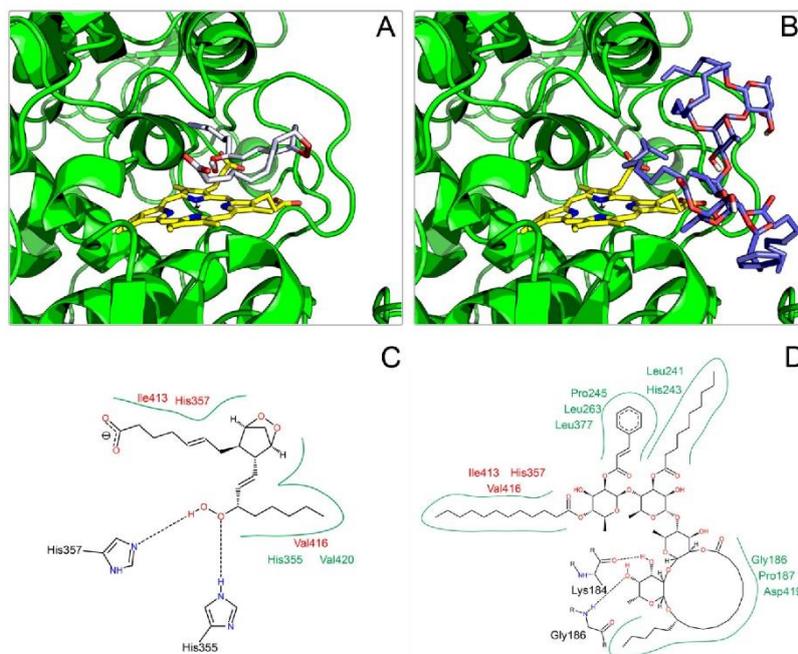
* ROESY contacts were obtained from previously published NMR Data [1].

2.3. Docking on COX

Since ipomotaoside A binding to COXs is not yet determined [1], docking calculations were performed on the enzymes' two catalytic sites, pointing to a better binding to the peroxidase cleft. The binding energy for ipomotaoside on cyclooxygenase site was extremely positive, suggesting a lack of binding to this region. This result reinforces the originality of this resin glycosides scaffold, since the activity of most non-steroidal anti-inflammatory drugs in therapeutic use modulates the cyclooxygenase site. Also, as no crystallographic structure of COXs is available complexed to compounds modulating the peroxidase site, its biological substrate, PGG₂, was docked for comparison and validation.

Accordingly, PGG₂ (Figure 4 and Supporting Information, Figure S7) demonstrated an orientation at peroxidase site in agreement with the catalytic mechanism of the enzyme, interacting directly with heme through PGG₂ hydroperoxide. Such an orientation is essential for PGG₂ to be reduced to PGH₂, and exert its role on inflammatory processes [6], reinforcing the adequacy of the employed docking proceedings.

Figure 4. Complexes obtained for PGG₂ (A,C) and ipomotaoside 1 (B,D) with COX-1 as derived from docking calculations. Heme group is highlighted in yellow. 2D images, generated with the PoseView server [17], present only interactions between the ligand and amino acid residues. Common amino acid residues between PGG₂-COX-1 are indicated in red.



Like PGG₂, ipomotaoside A demonstrated a direct interaction between heme and an acyl chain on both COX-1 and COX-2 (Figure 4 and Supporting Information, Figure S7), suggesting that this acyl group is the main responsible for the pharmacological effect of ipomotaoside A. Additionally, hydrophobic amino acids residues from peroxidase site, participating in PGG₂binding, also interact with ipomotaoside (Figure 4 and Supporting Information, Figure S7), emphasizing the similarity of both molecules' mechanism of interaction with the target enzyme.

3. Experimental

3.1. Nomenclature, Topologies and Software

The recommendations and symbols of nomenclature as proposed by IUPAC [18] were used. The relative orientation of a pair of contiguous carbohydrate residues was described by two torsional angles at the glycosidic linkage, Φ and Ψ , for α -L-Rha-(1 \rightarrow 2)- β -D-Fuc and α -L-Rha-(1 \rightarrow 4)- α -L-Rha, as shown below:

- α 1 \rightarrow 2:
 $\Phi = \text{O5} - \text{C1} - \text{O1} - \text{C2}'$

$$\Psi = C1 - O1 - C2' - C1'$$

- $\alpha 1 \rightarrow 4$:

$$\Phi = O5 - C1 - O1 - C4'$$

$$\Psi = C1 - O1 - C4' - C3'$$

Initially, the disaccharide structures were built using MOLDEN, [19] and the topologies of saccharide residues and ipomotaosides were generated by the PRODRG server [20]. Structures were manipulated using PyMOL [21]. Simulations and analyses were performed using GROMACS [22] simulation suite, version 4.5.1, and the GROMOS96 43a1 [23] force field, while metadynamics calculations [24] were performed using a modified version of GROMACS 4.5.1 interfaced with the PLUMED plugin package, version 1.2.2 [25]. The free energy surfaces were obtained through *sum_hills* tool from PLUMED package.

3.2. Topology Construction

Each ipomotaoside was built with the most prevalent conformations of its minimal components in solution (disaccharides). All compounds were built using MOLDEN and submitted to the PRODRG server to receive their crude topologies and atomic coordinates. HF/6-31**^{††}-derived Löwdin atomic charges, from previous studies [26], and improper dihedrals for maintenance of the conformational states of the monosaccharides α -L-rhamnose (1C_4), and β -D-fucose (4C_1), were included. Additionally, proper dihedral angles were included according to the GROMOS96 force field. Pyridine solution values were retrieved from the literature [12].

3.3. Metadynamics

Metadynamics calculations for the isolated disaccharides 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 θ and φ angular coordinates of Cremer and Pople [27].

3.4. MD Simulations

Each minimum energy conformation of disaccharides, obtained from the metadynamics, as well as the complete ipomotaosides, was submitted to MD simulations in pyridine and water (SPC water model) [28] in a triclinic box using periodic boundary conditions. Counterions (Na^+) were added to neutralize the system charge when necessary. The systems were submitted to energy minimization by steepest Descents algorithm and subsequently to MD simulations. The Lincs method [29] was applied to constrain covalent bond lengths, allowing an integration step of 2 fs. The particle mesh Ewald method [30] was applied in the calculation of electrostatic interactions. Temperature and pressure were kept constant by coupling ipomotaosides (or carbohydrates), ions, and solvent to external temperature and pressure baths, with coupling constants of $\tau = 0.1$ and 0.5 ps, respectively. Finally, the simulations were performed at the constant temperature of 310 K for 0.1 μs .

3.5. ROESY Signals

ROESY $^1\text{H-NMR}$ data for compounds 1–4 [1] was employed to validate the intra-molecular H-H contacts observed on the performed simulations. The present work is based on a united-atom force field, which significantly reduce the computational costs, thus allowing faster simulations with longer time scales. Thus, to allow a comparison of the simulations to ROESY data, nonpolar hydrogens atoms were added to frames retrieved from trajectories, at every 10 ps, for each ipomotaoside. The correct geometry and hybridization were respected, and the obtained models were used to calculate the average interproton distances from simulations.

3.6. Docking Procedures

Docking calculations were performed with Autodock, version 4.2 [31]. COXs from PDB IDs 1Q4G and 1CVU were used. PGG₂ was built as described for ipomotaosides. Water molecules were removed prior to docking procedures. The Lamarckian Genetic Algorithm (LGA) was used to explore the binding sites. For each run, a maximum number of energy evaluations was set to 250,000,000, and a maximum number of 27,000 LGA operations was generated on populations of 10 individuals. In the current work 100 runs were performed, in a population of 1,000 individuals for each ligand. Crossover, mutation, and elitism were set to 0.80, 0.02, and 1, respectively. All rotatable dihedral angles of hydrophobic chains were treated as flexible, while Φ and Ψ dihedral angles of disaccharides were maintained rigid in the most abundant conformational states. The orientations of ipomotaosides and PGG₂ at the binding site were selected from docked conformations as representative of the lower energy clusters generated by Autodock. The bidimensional images were generated with the PoseView server [17].

4. Conclusions

In the present study, ipomotaosides A to D had their tridimensional structures obtained from the most abundant conformational states of their glycosidic linkages. These compounds were submitted to refinement under MD simulations in both aqueous and nonaqueous solvents and were compared to experimental data. A minor influence of pyridine on the compounds dynamic was observed, which reinforces the validity of the previously described NMR data for the understanding of ipomotaosides' behavior under biological conditions. The reproduction of interproton ROESY contacts during unrestrained simulations on pyridine solvent suggests a precise conformational characterization of these flexible molecules, offering solution representative geometries for docking calculations. The so- obtained complexes pointed to ipomotaoside A as an inhibitor of the COX enzymes' peroxidase site, in a similar binding to that of PGG₂, providing a structural basis for future studies aiming at optimization of ipomotaosides as potential new anti-inflammatory agents.

Supplementary Materials

Supplementary materials can be accessed at: <http://www.mdpi.com/1420-3049/19/4/5421/s1>.

Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), MCT, by 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).

Author Contributions

H. V. conceived the computational study. P. R. A. developed the atomic models of ipomotaosides and performed molecular dynamics and analysis. L. G. S. and P. R. A. performed the docking calculations and analysis. C. S. G. and P. R. A. performed the analysis of ¹H NMR Spectroscopic Data of compounds 1–4. All authors read and approved the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Yoshikawa, K.; Yagi, C.; Hama, H.; Tanaka, M.; Arihara, S.; Hashimoto, T. Ipomotaosides A–D, resin glycosides from the aerial parts of *Ipomoea batatas* and their inhibitory activity against COX-1 and COX-2. *J. Nat. Prod.* **2010**, *73*, 1763–1766.
2. Li, S.Z. *Min Dynasty, Compendium of Materia Medica*; Medical Publishing House: Beijing, China, 1999; p. 1501.
3. Corona-Castañeda, B.; Pereda-Miranda, R. Morning glory resin glycosides as modulators of antibiotic activity in multidrug-resistant Gram-negative bacteria. *Planta Med.* **2012**, *78*, 128–131.
4. Figueroa-González, G.; Jacobo-Herrera, N.; Zentella-Dehesa, A.; Pereda-Miranda, R. Reversal of multidrug resistance by morning glory resin glycosides in human breast cancer cells. *J. Nat. Prod.* **2012**, *75*, 93–97.
5. Pereda-Miranda, R.; Rosas-Ramírez, D.; Castañeda-Gómez, J. Resin glycosides from the morning glory family. In *Progress in the Chemistry of Natural Products*; Kinghorn, D.A., Falk, H., Kobayashi, J., Eds.; Springer: New York, NY, USA, 2010; Volume 92, pp. 77–153.
6. Smith, W.L.; Garavito, R.M.; DeWitt, D.L. Prostaglandin Endoperoxide H Synthases (Cyclooxygenases)-1 and -2. *J. Biol. Chem.* **1996**, *271*, 33157–33160.
7. Rencurosi, A.; Mitchell, E.P.; Cioci, G.; Pérez, S.; Pereda-Miranda, R.; Imberty, A. Crystal Structure of Tricolorin A: Molecular Rationale for the Biological Properties of Resin Glycosides Found in Some Mexican Herbal Remedies. *Angew. Chem. Int. Ed.* **2004**, *43*, 5918–5922.
8. Woods, R.J. Computational carbohydrate chemistry: What theoretical methods can tell us. *Glycoconj. J.* **1998**, *15*, 209–216.
9. Dwek, R.A. Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* **1996**, *96*, 683–720.
10. Pérez, S.; Mulloy, B. Prospects for Glycoinformatics. *Curr. Opin. Struct. Biol.* **2005**, *15*, 517–524.

11. Pol-Fachin, L.; Fernandes, C.L.; Verli, H. GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations. *Carbohydr. Res.* **2009**, *344*, 491–500.
12. Pedebos, C.; Pol-Fachin, L.; Verli, H. Unrestrained conformational characterization of *Stenocereus eruca* saponins in aqueous and nonaqueous solvents. *J. Nat. Prod.* **2012**, *75*, 1196–1200.
13. Pol-Fachin, L.; Serrato, R.V.; Verli, H. Solution conformation and dynamics of exopolysaccharides from *Burkholderia* species. *Carbohydr. Res.* **2010**, *345*, 1922–1931.
14. Castro, M.O.; Pomin, V.H.; Santos, L.L.; Vilela-Silva, A.C.E.S.; Hirohashi, N.; Pol-Fachin, L.; Verli, H.; Mourao, P.A.S. A Unique 2-Sulfated β -Galactan from the Egg Jelly of the Sea Urchin *Glyptocidaris crenularis* Conformation Flexibility Versus Induction of the Sperm Acrosome Reaction. *J. Biol. Chem.* **2009**, *284*, 18790–18800.
15. Pol-Fachin, L.; Verli, H. Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Carbohydr. Res.* **2008**, *343*, 1435–1445.
16. Jiang, Z.; Geyer, A.; Schmitd, R.R. The Macrolidic Glycolipid Calonyctin A, a Plant Growth Regulator: Synthesis, Structural Assignment, and Conformational Analysis in Micellar Solution. *Angew. Chem. Int. Ed.* **1995**, *34*, 2520–2524.
17. Stierand, K.; Maaß, P.; Rarey, M. Molecular complexes at a glance: Automated generation of two-dimensional complex diagrams. *Bioinformatics* **2006**, *22*, 1710–1716.
18. Horton, D. Nomenclature of carbohydrates. *Pure Appl. Chem.* **1996**, *68*, 1919–2008.
19. Schaftenaar, G.; Noordik, J.H.J. Molden: A pre- and post-processing program for molecular and electronic structures. *Comput.-Aided Mol. Des.* **2000**, *14*, 123–134.
20. Schuettelkopf, A.W.; van Aalten, D.M.F. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* **2004**, *D60*, 1355–1363.
21. DeLano, W.L. *The PyMOL Molecular Graphics System*; DeLano Scientific LCC: San Carlos, CA, USA, 2002.
22. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
23. Scott, W.R.P.; Hünenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Krüger, P.; van Gunsteren, W.F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
24. Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843.
25. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raitieri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.A.; *et al.* PLUMED: A portable plugin for free energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
26. Verli, H.; Guimarães, J.A.A. Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution. *Carbohydr. Res.* **2004**, *339*, 281–290.
27. Cremer, D.; Pople, J.A. A General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
28. Berendsen, H.J.C.; Grigera, J.R.; Straatsma, T.P.J. The missing term in effective pair potentials. *Phys. Chem.* **1987**, *91*, 6269–6271.
29. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

30. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
31. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comp. Chem.* **2009**, *16*, 2785–2791.

Sample Availability: Not available.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

10. CURRICULUM VITAE

I. Formação acadêmica:

Mestrado em Biologia Celular e Molecular pela Universidade Federal do Rio Grande do Sul, de 2009 a 2011. *Caracterização estrutural e conformacional de Prostaglandina Endoperóxido Sintases*. Orientador: Prof. Hugo Verli.

Graduação em Farmácia pela Universidade Federal do Rio Grande do Sul, de 2004 a 2008. *Glicosilação de proteínas*. Orientador: Prof. Hugo Verli.

II. Formação complementar: (restrito ao período de desenvolvimento da tese)

Hands on Course: Coarse Grain Methods for Biomolecular Simulations (Curso de curta duração), Institut Pasteur, Montevideo, Uruguai.

Métodos de Docking Receptor-Ligante, (Curso de curta duração), Laboratório Nacional de Computação Científica, Dr. Laurent Dardenne, Dra. Camila Magalhães, 2010.

Simulação de Proteínas de Biomembranas, (Curso de curta duração), Laboratório Nacional de Computação Científica, Dr. Hubert Stassen, Dr. Rafael Bernardi.

III. Atuação Profissional: (restrito ao período de desenvolvimento da tese)

Universidade de Santa Cruz do Sul, UNISC, Brasil. Vínculo institucional: Celetista, Enquadramento Funcional: Professor, Carga horária: 4h/semana.

Disciplina: Bioquímica dos alimentos

Departamento de Biologia e Farmácia

IV. Trabalhos científicos apresentados em congressos: (restrito ao período de desenvolvimento da tese)

Ligabue-Braun, R; Sachett, L.G.; Carlini, CR; Verli, H. INTERACTION OF THE MAJOR CAT ALLERGEN (FEL D 1) WITH CALCIUM. XLII Annual Meeting of SBBq, Foz do Iguaçu, 2013.

Sachett, L.G.; Verli, H. STRUCTURAL AND DYNAMICAL CHARACTERIZATION OF AN α -1,6-FUCOSYLTRANSFERASE. In: XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012, Foz do Iguaçu – PR. XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq, 2012

Arantes, P.R.; Sachett, L.G.; Fernandes, C.L.; Graebin, C.S.; Verli, H. CONFORMATIONAL CHARACTERIZATION OF IPOMOTAOSIDES. In: XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012, Foz do Iguaçu – PR. XLI Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq, 2012.

VI. Publicações em periódicos especializados:

Fernandes, C.L.; Sachett, L.G.; Pol-Fachin, L.; Verli, H. GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins. Carbohydrate Research (Chicago, Ill. Print), v.345, p.663 - 671, 2010.

Sachett, L.G.; Verli, H. Dynamics of Different Arachidonic Acid Orientations bound to Prostaglandin Endoperoxide Synthases. European Journal of Medicinal Chemistry, v.46, 5212-5217, 2011.

Sachett, L.G.; Verli, H. O papel de carboidratos na função proteica: o caso das PGHS. Revista Virtual de Química, 2011, 3(4), 275-285.

Arantes, Pablo; Sachett, Liana; Graebin, Cedric; Verli, Hugo. Conformational Characterization of Ipomotaosides and Their Recognition by COX-1 and 2. *Molecules* (Basel. Online), v. 19, p. 5421-5433, 2014.

V. Educação e popularização de ciência e tecnologia:

Co-organização, Curso de Férias PPGBCM, "Plantas... Para que mesmo?", público-alvo: professores de ciências e alunos de ensino médio, 2011.

Co-organização, Curso de Férias PPGBCM, "Microorganismos... mocinhos ou bandidos?", público-alvo: professores de ciências e alunos de ensino médio, 2009.